



저작자표시-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

치의과학박사 학위논문

노인 만성질환자의 건강관련 삶의 질
영향요인 분석 및 예측모델 개발

2013년 2월

서울대학교 대학원
치의과학과 의료경영과정보학
이 수 경

노인 만성질환자의 건강관련 삶의 질
영향요인 분석 및 예측모델 개발

지도교수 김 정 은

이 논문을 치의과학박사 학위논문으로 제출함

2012년 12월

서울대학교 대학원

치의과학과 의료경영과 정보학

이 수 경

이수경의 치의과학박사 학위논문을 인준함

2012년 12월

위 원 장 이 재 일



부위원장 김 정 은



위 원 김 홍 기



위 원 손 연 정



위 원 강 보 영



Analysis of Influencing Factors and Development of
Prediction Model for the Health related Quality of Life
in Elderly with Chronic Diseases

by
Lee, Soo-Kyoung

A thesis submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy
in Healthcare Management and Informatics

Seoul National University, Seoul, Korea
December, 2012

Doctoral Thesis Committee:

Professor	<u>Lee, Jee-Il</u>	Chairperson
Professor	<u>Kim, Jeong-Eun</u>	Vice Chairperson
Professor	<u>Kim, Hong-Gee</u>	
Professor	<u>Son, Youn-Jung</u>	
Professor	<u>Kang, Bo-Yeong</u>	

국문 초록

인구 고령화와 이로 인한 만성질환의 증가는 결과적으로 노인들의 건강 관련 삶의 질(Health Related Quality of Life: HRQoL)을 저하시키는 원인이 되고 있다. 따라서 노인 만성질환자의 HRQoL에 영향을 미치는 요인을 파악하고, HRQoL이 저하된 위험군을 예측하여 HRQoL 향상을 위한 효율적인 중재방안을 제시할 필요가 있다.

이에 본 연구는 2008~2010년 국민건강영양조사(Korea National Health and Nutrition Examination Survey: KNHANES)라는 대규모의 광범위 조사자료를 이용하여 만성질환을 가진 노인을 대상으로 HRQoL에 영향을 미치는 요인들을 포괄적으로 분석하고, 중재가 필요한 HRQoL 위험군을 예측하는 모델을 개발하고자 시도되었다.

HRQoL에 관한 다각적인 분석을 위해서는 상당한 노력과 전문성이 요구되기에 이러한 복잡한 분석과정을 자동으로 지원해줄 수 있는 다양한 방법들을 적용할 필요가 있다. 특히 대용량 자료의 분석 및 예측에 용이한 기계학습(Machine Learning: ML) 기법을 새롭게 적용해 봄으로써 노인 만성질환자의 HRQoL 분석 및 예측을 위한 ML 기법의 적용 가능성과 그에 따른 문제점 및 해결방안을 함께 살펴보았다.

자료분석 및 예측모델 개발을 위해 SPSS (Statistical Package for the Social Sciences) version 20.0과 MATLAB (Matrix Laboratory) Revised for Version 7.14 Release 2012a 프로그램을 사용하여 기술통계, X^2 -test, 단계적 로지스틱 회귀분석(Stepwise Logistic Regression: SLR)을 실시하였으며, ML 기법 중 의사결정 나무(Decision Tree: DT), 랜덤 포레스트(Random Forest: RF), 지지벡터기계(Support Vector Machine: SVM) 방법을 적용하였다. HRQoL은 EQ-5D (European Quality of Life-5 Dimensions) 항목으로 측정되었으며, EQ-5D index 0.678을 기준으로 0.678 이상인 그룹(n=648)과 0.678 미만인 그룹(n=68)으로 분류하였다.

연구결과 노인 만성질환자의 HRQoL에 영향을 미치는 요인들로 ‘소득’, ‘만성질환 종류’, ‘우울’, ‘불편감’, ‘주관적인 건강상태’가 통계적으로 유의하였다. ML 적용결과 노인 만성질환자의 HRQoL 예측을 위한 추가적인 중요변수로서 ‘골관절염 소요기간’, ‘골다공증 유병여부’, ‘뇌졸중 진단시기’, ‘하악보철물 상태’, ‘1년간 자살생각여부’ 등이 도출되었다.

구축된 예측모델의 타당도 검증은 10 fold cross-validation 방법을 적용하였다. 성능평가를 위해서 Accuracy, F-score 등의 값을 모형별로 산출하여 비교한 결과, SLR이 Accuracy=0.93, F-score=0.49로 가장 우수한 성능을 나타내었으며, 그 다음으로는 SVM (0.90, 0.51), RF (0.87, 0.33), DT (0.82, 0.23) 순이었다.

결과적으로 KNHANES와 같은 혼합형태의 자료에 기계학습 기법을 적용함에 있어서 변수 선정과 자료 정규화(Data Normalization) 과정을 통해 예측모델의 성능과 효율을 향상시킬 수 있었다. 향후 이에 대한 지속적인 연구와 검증 과정이 요구되며, 유사형태의 자료를 활용하는 경우에 변수 선정과 자료를 정규화하는 방법에 대한 구체적인 가이드라인을 제시할 수 있을 것이다.

본 연구를 통해 만성질환을 가진 노인들의 HRQoL을 증진시키기 위한 맞춤형 건강관리 전략 및 중재 프로그램 개발에 요구되는 다양한 기초자료를 제공할 수 있을 것이다. 본 연구와 같이 의료분야에서의 자료 재창출(Data Repositioning)을 통해 가치있는 새로운 지식을 도출할 수 있도록 이러한 접근 방법이 더욱 활발하게 적용되기를 기대한다.

주요어 : 노인, 만성질환, 건강관련 삶의 질, 예측모델, 기계학습

학 번 : 2008-31047

목 차

국문초록	i
I. 서론	1
1. 연구 필요성	1
2. 연구 목적	3
3. 용어 정의	4
3.1 만성질환	4
3.2 건강관련 삶의 질	4
3.3 기계학습 기법	5
II. 문헌고찰	6
1. 인구 고령화에 따른 노인 만성질환자의 증가	6
2. 건강관련 삶의 질 개념 및 측정도구	9
3. 건강관련 삶의 질 개념적 모형	13
3.1 Stuifbergen의 만성질환자를 위한 건강증진 및 삶의 질 모형	13
3.2 Wilson과 Cleary의 질병과 삶의 질 연계모형	15
4. 노인 만성질환자의 건강관련 삶의 질 영향요인	18
4.1 개인적 특성과 환경적 특성	20
4.2 생리적 요인	22
4.3 증상경험	23
4.4 일반적 건강 지각	24
4.5 건강증진행위	25
5. 보건의료 분야에서의 기계학습 적용	26
5.1 의사결정나무(Decision Tree)	28

5.2 랜덤 포레스트(Random Forest)	29
5.3 지지벡터기계(Support Vector Machine)	31
III. 개념적 기틀 및 연구가설	34
1. 개념적 기틀	34
2. 가설적 모형	37
3. 연구가설	39
IV. 연구 방법	40
1. 연구 단계	40
2. 연구 대상	41
3. 연구 도구	43
3.1 EQ-5D index	43
3.2 개인적 특성	45
3.3 환경적 특성	45
3.4 생리적 요인	46
3.5 증상경험	47
3.6 일반적 건강 지각	48
3.7 건강증진 행위	48
4. 자료 정규화	49
4.1 Z-변형(Z-transformation)	49
4.2 이진화(1-of-K encoding scheme coding)	49
4.3 이산화(Discretization)	50
4.4 새로운 변수 생성	50
5. 자료 분석	52

V. 연구 결과	54
1. 대상자의 일반적 특성	54
2. 대상자의 건강관련 삶의 질 수준	57
3. 영향요인 분석	62
3.1 DecisionTree model	62
3.2 Random Forest model	65
3.3 Support Vector Machine model	67
3.4 Stepwise Logistic Regression model	67
3.5 Comparison of influencing factors	69
4. 가설모형 수정	74
5. 연구가설 검정	76
6. 예측모델 개발	78
6.1 DecisionTree model	78
6.2 Random Forest model	79
6.3 Support Vector Machine model	80
6.4 Stepwise Logistic Regression model	82
7. 예측모델 평가	83
7.1 DecisionTree model	83
7.2 Random Forest model	85
7.3. Support Vector Machine model	87
7.4. Stepwise Logistic Regression model	89
7.5 Comparison of all prediction models	91
8. 기계학습 기법의 적용가능성 탐색	93
VI. 논의	99
1. 연구결과에 대한 논의	99
2. 연구방법에 대한 논의	105

3. 연구의 활용 방안	107
3.1 학문적 의의	107
3.2 실무에의 적용	108
 VII. 결론 및 제언	111
1. 결론	111
2. 제언	112
 References	115
 Appendix	133
 Abstract	187

List of Tables

Table 1. Measurement tool of health related quality of life	12
Table 2. Influencing factors of health related quality of life	20
Table 3. Applications and case of machine learning	27
Table 4. Method table of data normalization(sample)	51
Table 5. Case selection result(unit: person)	54
Table 6. EQ-5D index by chronic disease groups	55
Table 7. Descriptive statistics for observed variables	56
Table 8. EQ-5D index by groups	58
Table 9. Differences of health-related quality of life by general characteristics	59
Table 10. Important variables in DT	64
Table 11. Heat map of important feature variables in RF	65
Table 12. Result of stepwise logistic regression	68
Table 13. Important variables of four models	69
Table 14. Mean score of EQ-5D index by influencing factors of ML	72
Table 15. Confusion matrix of DT	83
Table 16. Confusion matrix of RF	85
Table 17. Confusion matrix of SVM	87
Table 18. Confusion matrix of SLR	89
Table 19. Comparison of performance in prediction models	91
Table 20. Re-examination sets	93
Table 21. Variables of re-examination sets	94
Table 22. Variables of hierarchical clustering set	96
Table 23. Result of re-examination	97
Table 24. Survey progress of KNHANES	134
Table 25. Survey contents of KNHANES	135

List of Figures

Figure 1. Survey of health behavior	8
Figure 2. Conceptual model of health promotion and quality of life for people with chronic and disabling conditions	14
Figure 3. Conceptual model of link between disease and quality of life	16
Figure 4. Revised Wilson and Cleary model for health-related quality of life	17
Figure 5. A general architecture of a random forest (Verikas, Gelzinis, & Bacauskiene, 2011)	30
Figure 6. Pattern classification by an SVM(Takahashi, 2012)	31
Figure 7. Construction of hypothetical model	35
Figure 8. Conceptual framework of prediction model of health related quality of life in elderly with chronic diseases	36
Figure 9. Hypothetical model of health related quality of life in elderly with chronic diseases	37
Figure 10. Research process	40
Figure 11. Procedure for extracting data and analysis	42
Figure 12. Distribution of EQ-5D index	58
Figure 13. Analysis result of decision tree	63
Figure 14. Modified conceptual framework of prediction model of health related quality of life in elderly with chronic diseases	74
Figure 15. Decision trees graph	78
Figure 16. F-score by number of trees	79

Figure 17. Feature importance	80
Figure 18. Performance per number of features in SVM	81
Figure 19. Performance contour with parameter c & γ in SVM	81
Figure 20. Boxplot of performance in DT	84
Figure 21. Boxplot of performance in RF	86
Figure 22. Boxplot of performance in SVM	88
Figure 23. Boxplot of performance in SLR	90
Figure 24. Rader chart of performance of four models	92
Figure 25. Dendrogram of hierarchical clustering	95
Figure 26. Proposed system architecture	109
Figure 27. Virtuous cycle of data repositioning	114
Figure 28. History of KNHANES implementation	133
Figure 29. Survey methods of KNHANES	135

List of Appendices

Appendix 1. Survey of KNHANES	133
Appendix 2. Survey table of EQ-5D	136
Appendix 3. Weighted estimate of EQ-5D 243 cases	137
Appendix 4. Method table of variables normalization	140
Appendix 5. MATLAB code of DT, RF, SVM algorithm	168

I. 서론

1. 연구 필요성

우리나라는 2011년 기준 65세 이상 노인인구가 5,184천명(전체 인구의 10.5%)으로 이미 고령화 사회로 진입하였고, 65세 이상 노인 진료비의 경우 15조 3,768억원으로 전체 진료비의 33.3%에 해당되며, 지속적인 증가 추세에 있다(건강보험심사평가원, 2012). 이처럼 매년 증가하는 노인 진료비는 중요한 사회적 이슈가 되었으며, 이를 해결하기 위한 다양한 방안들이 요구되고 있다. 또한 노인들의 만성질환은 완치를 목적으로 하는 것이 아닌 관리 및 유지를 통해 건강관련 삶의 질(Health Related Quality of Life: 이하 HRQoL)을 높이는데 있으며, HRQoL은 노인 만성질환의 치료효과 및 사망률, 의료비용 등에 영향을 미치는 중요한 문제로 대두되고 있다.

HRQoL이란 건강수준 또는 질병정도에 의해 영향을 받는 삶의 질을 일컫는 말로 일상생활, 주관적 경험에 관한 신체적, 사회적, 신체 감각적 그리고 주관적 행복을 포괄한 광범위하고 다면적인 개념(Kempen, Ormel, Brilman, & Relyveld, 1997)이다. 이러한 복잡한 개념을 구체적으로 이해하고 접근하기 위해서는 다각도의 분석이 요구되며, 기존 분석방법과 상호보완적인 새로운 분석방법을 도입하여 다양한 영향요인을 탐색하고 분석함으로써 당면하고 있는 문제를 보다 효과적으로 해결할 필요가 있다.

HRQoL관련 연구로는 대상을 성인(길선령 et al., 2008; 안홍모, 2010), 여성(김미령, 2006; 손연정, 2007) 등 폭넓게 접근하거나 기러기아빠(차은정, 2012), 기혼 직장여성(손연정, 2007), 취약계층 남성노인(전은영 & 최연희, 2010) 등 특정 대상과 당뇨(Porojan, Poanta, & Dumitrascu, 2012; 최해림, 2008), 뇌졸중(Alguren, Fridlund, Cieza, Sunnerhagen, & Christensson, 2012), 뇌성마비(이병희 & 고주연, 2010) 등 특정 질환으로 한정하여 접근하기도 하였다.

연구방법론에 있어서는 주로 설문 기반의 한정된 자료만을 분석하였고, 대부분 전통적인 통계방법인 회귀분석, 구조모형 분석(Heckman, 2003; 장선주, 2010; 차은정, 2012) 방법을 사용하였으며, 의사결정 나무(Decision Tree: 이하 DT)(최선영, 2009)의 단일한 방법만을 적용하여 제한적으로 분석한 연구도 있었다.

만성질환을 가진 노인들을 대상으로 한 HRQoL에 관한 연구들을 살펴보면 남해성(2000)은 도시지역 노인의 경우 신체거동과 통증과 관련된 관절 질환, 뇌졸중, 편두통 등의 질환과 높은 관련성을 보였다고 보고하였으며, 이동호(2011)는 근골격계 질환이 HRQoL이 가장 낮았으며, 주관적 건강상태, 근골격계 질환수, 의학적 건강상태 순으로 영향을 미친다고 하였고, 남상권(2011)은 근골격계 질환과 복합된 질환의 경우 낮은 점수를 보였고, 심리적 상태, 신체적 건강, 소득 순으로 영향력이 높았다고 보고하였다. 이처럼 노인 만성질환자들을 대상으로는 만성질환의 유형 또는 기간 등 일부 내용만을 다루는 제한적인 연구들이 시행되었을 뿐이다.

이에 본 연구는 검진자료를 포함한 대규모의 광범위 국가조사자료를 활용하여 노인들의 만성질환 유형뿐만 아니라 만성질환 개수와 기간을 포함한 다양한 영향요인들에 관하여 보다 포괄적으로 분석하고, 중재가 필요한 HRQoL 위험군을 예측하는 다양한 모델을 개발하고자 한다.

HRQoL에 관한 다각적인 분석을 위해서는 상당한 노력과 전문성이 요구되기에 이러한 복잡한 분석과정을 자동으로 지원해줄 수 있는 다양한 방법들을 적용할 필요가 있다. 특히 대용량 자료의 분석 및 예측에 용이한 기계학습(Machine Learning: 이하 ML) 기법을 새롭게 적용해 봄으로써 노인 만성질환자의 HRQoL 분석 및 예측을 위한 ML 기법의 적용 가능성과 그에 따른 문제점 및 해결방안을 함께 살펴보고자 한다.

대용량의 방대한 데이터가 빠른 속도로 전송되는 스마트 환경에서 빅데이터 기술의 부상은 필연적이다(손민선 & 문병순, 2012). 최근 빅데이터 시대를 맞이하여 이를 분석하여 지능형 서비스로 활용할 수 있는 기술로 인공지능 기술이 다시 관심을 받고 있으며, 이를 활용하기 위한 분석기술로

인공지능의 여러 요소 기술 중 ML 기법이 매우 유용하게 사용되고 있다(임수중, 민옥기, & Sw원천기술연구팀, 2012).

통계적 방법론으로는 로지스틱 회귀분석(Logistic Regression: 이하 LR)이 전통적으로 이용의 편의성과 변수조절의 간편성 때문에 주로 사용되나 데이터에 따라 생성되는 분류기(classifier)가 다양하여 불안정하다는 단점을 지니며(정승일, 2010), 이외에 패턴 분류의 한 도구로서 최근 광범위하게 연구되고 있는 ML 기법 중에서 감독 학습(Supervised Learning), 즉 분류(Classification) 기법들을 다양하게 적용해 볼 수 있을 것이다.

본 연구에서는 전통적인 방법인 LR과 ML 기법 중 분류에 적합한 DT, 랜덤 포레스트(Random Forest: 이하 RF), 지지벡터기계(Support Vector Machine: 이하 SVM) 방법을 적용하여 노인 만성질환자의 HRQoL에 관한 영향요인을 포괄적으로 분석하고, HRQoL 위험군을 예측하는 최적의 모델을 개발함으로써 HRQoL을 향상시킬 수 있는 효율적인 중재 방안을 제시하고자 하였다.

2. 연구 목적

본 연구는 2008~2010년 국민건강영양조사(Korea National Health and Nutrition Examination Survey: 이하 KNHANES)라는 대규모 조사자료를 이용하여 만성질환을 가진 노인을 대상으로 HRQoL에 영향을 미치는 요인들을 파악하고, HRQoL 위험군을 예측하는 모델을 개발하고자 하였다. 구체적인 목적은 다음과 같다.

첫째, 노인 만성질환자의 HRQoL에 영향을 미치는 요인을 탐색한다.

둘째, 노인 만성질환자의 HRQoL 위험군 예측모델을 개발하고 평가한다.

셋째, 노인 만성질환자의 HRQoL 위험군 예측모델에 있어 기계학습 기법의 적용 가능성을 탐색한다.

3. 용어 정의

3.1 만성질환

만성질환이란 정상적인 생리적 기능이 서서히 진행적으로 쇠퇴하고 영구적인 장애를 초래하여 재활요법 및 장기간의 감독, 관찰, 간호 등이 요구되는 특징을 지닌 실체를 말한다(Berton, Cassel, & Gore, 1977).

본 연구에서는 KNHANES 조사항목 중 질병이환에 관한 설문에서 만성질환에 해당되는 고혈압, 고지혈증, 뇌졸중, 협심증, 심근경색증, 당뇨병, 갑상선 질환, 신부전, 골관절염, 류마티스성 관절염, 골다공증, 폐결핵, 천식, 위암, 간암, 대장암, 유방암, 자궁경부암, 폐암, 기타암, B형 간염, C형 간염, 간경변증만을 포함하였다. 또한 연구대상으로 포함하는 기준으로서 상기 만성질환의 유병여부 ‘있음’과 의사진단여부 ‘있음’으로 대답하였으며, 만성질환 진단시기가 최소 1년 이상 경과한 경우를 대상으로 하였다

3.2 건강관련 삶의 질

HRQoL은 신체적 건강과 기능, 정신적 건강, 사회적 기능과 일반적인 안녕의 영역을 포함하는 다차원적 개념이다(Marquis, Davies, & Ware, 1983). 본 연구에서는 KNHANES의 EQ-5D (European Quality of Life-5 Dimensions) 측정 결과를 이용하였다.

EQ-5D는 EuroQoL group이 건강관련 삶의 질 상태를 기술하고 가치화하기 위해 개발한 도구로 운동능력, 자기관리, 일상활동, 통증/불편감, 불안/우울의 5개 항목에 대해 ‘지장 없음, 다소 지장 있음, 매우 심함’의 3가지 수준으로 평가하도록 되어있다. EQ-5D index는 한국인 질 가중치 모형 연구(Nam, Kim, Kwon, Koh, & Poul, 2007)를 반영하여 산출된 지표로, 완전한 건강상태는 1이며 가장 불완전한 상태는 -0.171이다.

3.3 기계학습 기법

ML 기법이란 인간이 가지고 있는 고유의 지능적 기능 중 하나인 학습능력을 기계를 통해 구현하는 방법들에 대한 연구이다. 이 기법은 주어진 데이터들을 분석하여 그로부터 일반적인 규칙이나 새로운 지식을 자동적으로 추출해내는 방법론들을 개발하는 방법이다(박혜영 & 이관용, 2011).

본 연구에서는 ML 기법 중 DT, RF, SVM 방법을 적용하였다. DT는 보편적으로 많이 사용되며, RF와 SVM은 최근 새롭게 부각되고 있는 ML 기법으로서, 본 연구에서는 최적의 분류를 위한 예측모델을 개발하기 위해 3가지 방법을 모두 적용하여 비교하였다.

II. 문헌고찰

1. 인구 고령화에 따른 노인 만성질환자의 증가

노인에게는 노화과정과 더불어 건강문제들이 다양하게 나타나는 특성이 있으며, 노화로 인한 노인들의 건강특성으로는 평균 5~6가지가 중복적, 동시다발적으로 질병이 진행되는 경우가 많고, 다양한 신체적, 심리적인 변화를 가져온다(최영희, 2007). 노화는 피할 수 없는 변화과정으로 노인들은 이러한 과정에서 발생하는 많은 문제들을 효율적으로 대처하여야 하며, 이제는 개인의 문제가 아닌 사회 문제로 중요한 이슈가 되고 있는 만큼 체계적인 접근과 총체적인 해결방안을 모색할 필요가 있다.

만성질환은 만성적 과정을 가지는 질환으로 비전염성이며 퇴행성인 질환을 말하며, 일반적으로 불명확하고, 다원적인 위험요인, 긴 잠복기, 질병의 연장된 경과, 비전염성, 기능적 장애 또는 불구 그리고 치료되지 않는 특성을 가진다(보건산업진흥원, 2007). 또한 완치가 불가능한 질환으로 3개월 혹은 6개월 이상 질병이 지속되는 경우를 말한다(Summer, O'Neill, & Shirey, 1999). 만성질환은 질병의 원인이 뚜렷하게 밝혀지지 않고 여러 가지 복합요인에 의해 유발되고, 특히 생활양식과 관련이 높으며 일단 발병하면 치료가 어렵고 유병기간이 길어서 신체적, 정신적 고통과 치료를 위한 경제적 부담이 크다(Han, 2003).

노인의 건강문제는 대부분 만성질환으로 다양하면서도 단계적인 증상이 발생하고, 지속적이고 장기간에 걸쳐 건강상태는 점차 악화되어, 그 치료 및 관리가 매우 복잡한 양상을 지닌다. 이로 인한 신체기능의 제한이나 장애는 노인의 삶의 질을 저하시킬 뿐만 아니라 장기적인 치료를 요하기 때문에 의료에 대한 수용증가와 전체 의료비 증가에 커다란 요인으로도 작용하고 있다(남상권 & 심옥수, 2011).

보건복지부 2011년도 노인 실태조사(한국보건사회연구원, 2012)에서 전

국 11,542명의 노인을 면접 조사한 결과, 노인의 88.5%가 만성질환을 갖고 있고, 만성질환의 수는 평균 2.5개였으며, 3개 이상 지닌 경우도 44.3%에 달하였다. 60대는 15.5%, 70대는 35.9%, 85세 이상은 67.1%로 연령이 높아질수록 만성질환의 유병률이 높아졌으며, 고혈압이 54.8%로 가장 높았으며, 관절염(40.4%), 당뇨병(20.5%)의 순서로 유병률을 나타내었다.

만성질환 노인에 대한 기존 연구로는 건강상태 조사(문은영, 2004; 박은자, 2010; 정재임 & 김경아, 2002), 만성질환이 미치는 영향(최령, 문현주, & 함병덕, 2010), HRQoL(남상권 & 심옥수, 2011), 치료 중재 (김홍열 & 윤숙영, 2003; 송미순, 2004), 가족 부양자(이현지, 2007)에 이르기까지 다양한 주제를 다룬 연구가 진행되었다. 노인에서 삶의 질은 가족과 친구들을 포함한 사회적 지지나 개인 삶의 가치관을 높여주는 종교생활 그리고 경제 수준 등의 여러 요인이 있지만 그 중에 가장 큰 영향을 주는 요인은 건강문제다(Larson, 1978).

한국인의 만성질환에 대한 연구(성은주, 2005)에 따르면, 2000년 이후 한국인의 5대 사망원인은 모두 만성질환이었는데, 이는 대부분이 잘못된 생활습관이 원인이라고 지적하였다. 특히 만성질환은 노인의 삶의 질에 큰 영향을 미치는 것으로 생활습관의 개선이 건강장수를 가능케 함을 암시하였다. 농촌 독거노인의 건강상태와 삶의 질 연구(최연희, 2004)에서도 건강 유지 및 증진, 질병예방 등을 위해 개인적으로 행해지는 건강증진행위가 노인의 만성질환 유병률을 감소시키고, 수명을 연장하여 보다 건강한 노후를 보낼 수 있다고 제언하였다. 노인 만성질환자에게 필요한 태도로는 만성질환의 수용, 자신감의 함양이 필요하고, 자기관리 수행의 정도에 따라 건강상태, 의료비용, 생리적 지표 등이 달라지며 궁극적으로 삶의 질이 좌우된다(송미순, 2004).

보건복지부 2011년도 노인 실태조사(한국보건사회연구원, 2012)에서 노인들의 건강행태를 조사한 결과 흡연율은 12.6%, 연간 음주율은 33.8%, 운동 실천율은 50.3%이며, 지난 2년간 건강검진을 받은 노인은 81.6%였으며, 이는 2004년 조사 결과에 비해 흡연율은 감소하였으나, 음주율은 큰 변

화가 없고, 운동 실천율과 건강 검진율은 월등히 증가하였다(그림 1).

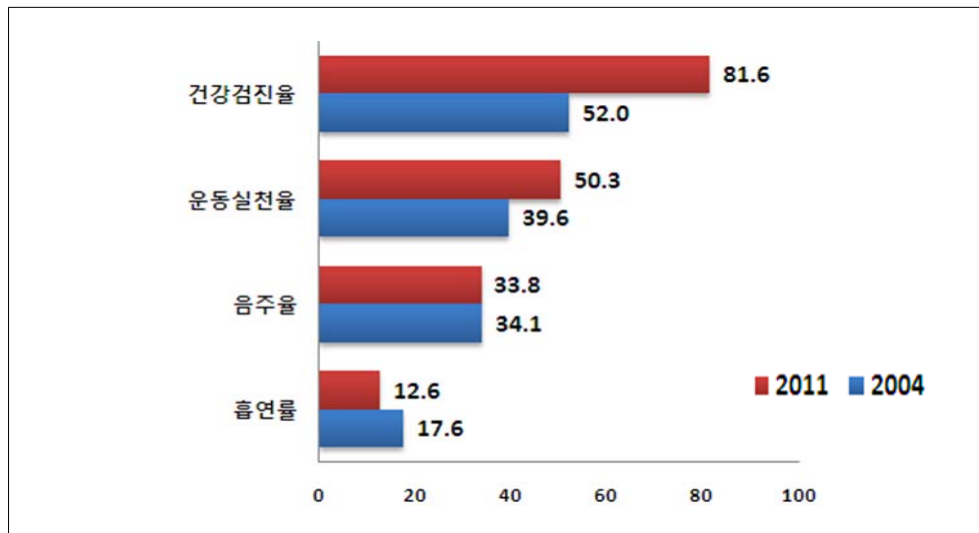


Figure 1. Survey of health behavior

이러한 노인 만성질환 관련 문제를 해결하기 위해서는 국가적으로 노인 만성질환자를 대상으로 생활습관을 개선하는 건강증진 행위와 정서적 함양을 도모하여야 한다. 특히 음주에 대한 건강행위 개선이 우선적으로 요구되며, 건강위험군에 대한 방문건강관리서비스 제공, 만성질환 예방을 위한 건강생활 실천 캠페인, 조기 검진 및 치료 관리 지원을 위한 국가차원의 지역사회 노인 만성질환 관리체계를 구축하고 강화할 필요가 있다.

인구의 고령화와 그로 인한 만성질환의 증가는 결과적으로 노인들의 건강상태 악화를 초래하며, 노인의 HRQoL 저하를 가져오는 원인이 되고 있다(고승덕 & 조숙행, 1997). 따라서 노인 만성질환자의 HRQoL에 관여하는 영향요인을 탐색하는 일은 실로 중요하다고 생각된다.

2. 건강관련 삶의 질 개념 및 측정도구

삶의 질은 건강관련 삶의 질(Health-related quality of life: HRQoL)과 건강과 관련되지 않은 삶의 질(Non-health-related quality of life: NHRQoL)로 구분할 수 있다. 건강관련 삶의 질이란 건강수준 또는 질병 정도에 의해 영향을 받는 삶의 질을 일컫는 말로 일상생활, 주관적 경험에 관한 신체적, 사회적, 감각적 그리고 주관적 행복을 포괄한 광범위하고 다면적인 개념이다(Kempen, et al., 1997). 건강과 관련되지 않은 삶의 질 요소로는 생활수준, 범죄율, 기후, 환경 등이 있고, 건강에 간접적으로 영향을 미치는 요소들에 의해 영향을 받는 삶의 질의 한 부분을 의미한다(Spilker, 1996).

HRQoL은 건강에 영향을 미치는 삶의 질을 나타내는 개념으로 일반적인 삶의 질과는 다른 개념이다. HRQoL은 매우 복잡하고 다면적인 개념으로서 그 이해를 돕기 위해 의료용어체계에서 quality of life 개념을 찾아보았다. 의학 자료의 주제를 나타내는 Medical Subject Headings (MeSH)에서는 삶의 질을 신체적, 정치적, 도덕적, 사회적 환경 등과 같은 삶의 속성의 수정 및 개선과 관련된 관심을 반영한 일반적인 개념, 즉 인간 삶의 질의 전체적인 상태로 정의하고 있다(NLM, 2012).

Ware, Kosinski, Dewey & Gandek (2001)은 HRQoL을 개인의 건강상태와 안녕보다는 더 포괄적인 개념으로서 전반적인 건강, 신체적 기능, 신체적 역할, 통증, 활력, 사회적 기능, 정신적 기능, 정서적 역할 등에서의 주관적인 평가를 의미한다고 하였다.

Wilson & Cleary (1995)는 HRQoL을 개인의 전반적인 삶의 행복도와 만족도에 연관된 주관적인 안녕이라고 정의하였고, Sajid, Tonsi & Baig (2008)는 질병의 정신사회적 영향을 파악하고자 하는 욕구에서 발전된 다차원적이고도 역동적인 개념으로서 경제적 안녕, 지역사회 환경의 특성과 건강상태를 포함한다고 정의하여 HRQoL의 다차원적인 속성을 보여주었다.

본 연구에서는 HRQoL에 대한 다양한 개념정의를 바탕으로 ‘노인 만성질

환자들 개개인의 삶을 구성하고 있는 다차원적 요인들에 대하여 개인이 느끼는 주관적 삶의 만족감'으로 정의하였다. 본 연구에서는 이러한 개념정의에 의해 측정이 가능하도록 노인 만성질환자들의 HRQoL에 미치는 영향요인들을 분석하고자 한다.

국내외적으로 만성질환자들을 위해 개발된 다양한 치료중재에 대한 효과를 평가하는 방법은 주로 객관적 결과를 사용하였으나, 최근에는 의료 소비자가 스스로 지각하는 자신의 사회적 기능과 역할, 증상, 신체적 기능 및 정서 상태와 같은 주관적 관점들도 같이 평가되어야 할 중요한 지표로 인식되기 시작하였다. 특히 이러한 주관적 결과지표들을 포괄하는 HRQoL을 중요한 평가지표로 함께 사용할 것을 권장하고 있다(이은현 et al., 2011).

HRQoL은 최근 여러 분야에서 중요한 개념으로 강조되고 있으며(Coons, Rao, Keininger, & Hays, 2000; 박상규, 2006; 이병희 & 고주연, 2010), 특히 노년기의 HRQoL을 향상시키기 위한 노력은 노인복지제도의 궁극적인 목표로서 노인의 연장된 삶을 보다 건강하고 만족스럽게 유지하는데 중점을 두고 있다. 따라서 만성질환을 가진 노인 등 특정 취약집단을 선별하여 HRQoL을 향상시키기 위해서는 맞춤형 중재방안을 위한 구체적인 전략이 요구된다.

국민개개인의 HRQoL 향상은 국가보건복지정책의 궁극적인 목적이며, 노령화 사회를 대비하기 위한 보건복지 정책분야의 핵심과제이다. 새로 도입된 보건정책의 효과를 평가하고, 진행 중인 각종 국가 보건의료서비스의 구조와 과정에 대한 평가과정에서 건강상의 효과뿐만 아니라 경제성을 고려한 다양한 평가를 시도하기 위해서는 결과지표로서 향상된 건강수준과 삶의 질이 측정되어야 한다(Nam, et al., 2007).

HRQoL 측정도구는 일반 인구집단을 대상으로 하는 일반 도구(General instrument)와 특정 질병을 가진 집단을 대상으로 하는 질병 특이적 도구(Disease-specific instrument)로 구별할 수 있다. 이 중 연구의 특성에 따라 포괄적인 도구를 사용할 지 아니면 구체적인 측정도구를 사용할 지를 결정하여야 한다(신상진, 2010). 질병 특이적 도구는 암, 간질, 천식, 요실금,

정신분열증, 비만 등 다양한 질병을 대상으로 하고 있다.

HRQoL 측정에 있어서 표 1과 같이 다양한 도구들이 사용되고 있으며, 이러한 도구들은 한국어판으로 번역이 되어 국내에서도 사용되고 있다. 이 중에서 만성질환 치료에 대한 측정도구인 FACIT(Functional Assessment of Chronic Illness Therapy Measurement System)은 개별적인 만성질환을 각각 별도로 다루고 있다.

Table 1. Measurement tool of health related quality of life

Type	Name	Method	Number of Question
General	COOP-C (COOP/WONCA chart)	Self, health care provider	6/8
	DUKE (Duke Health Profile)	Self, Interview	17
	EQ-5D (EuroQoL-5 Dimensions)	Phone, Self, Proxy	5
	HAQ (Health Assessment Questionnaire)	Phone, Self, Interview	22
	HUI (Health Utility Index)	Phone, Self, Proxy, Interview, Internet	15/40
Cancer	SF-36 (Medical Outcome Study Short Form-36)	Interview, Internet, Phone, Self	36
	EORTC QLQ C-30 (European Organization for Research and Treatment of Cancer Quality of Life Questionnaire)	Self	30
	Cancer caregiver CQOLC (Caregiver Quality of Life Index -Cancer)	Self	35
	Chronic disease therapy FACIT (Functional Assessment of Chronic Illness Therapy Measurement system)	Self, Interview	27
	Epilepsy QOLIE (Quality of Life in Epilepsy Inventory)	Self	10/31
Asthma	AQLQ (Asthma Quality of Life Questionnaire)	Self, Interview, Internet	32
Urinary Incontinence	ICIQ-UI Short Form (International Consultation on Incontinence Questionnaire-Urinary Incontinence Short Form)	Self	4
Obesity	OWLQOL (Obesity and Weight-Loss Quality of Life Instrument)	Self	17
Schizophrenia	SWN (Subjective Well-being under Neuroleptics scale)	Self	20/38

HRQoL을 측정하는 주요한 도구들의 타당성에 대한 선행연구(Coons, et al., 2000)에 따르면 어떤 우위를 갖는 도구는 없으며 대상자의 특징과 조사 환경을 고려하여 선정하도록 추천하고 있다. 이 중 EQ-5D는 아동에서 노인에 이르기까지 다양한 연령층을 대상으로 도구 타당성이 입증되었으며(Benedict et al., 2005), 개발 당시부터 7개국 언어로 번역되어 세계적으로 임상 및 역학조사 분야에서 민족과 국가 간 차이를 확인하는데 유용하였다(J Taylor, 2001).

우리나라에서도 HRQoL에 관한 중요성을 인식하고, 국가적으로 시행하고 있는 NHANES에서 2005년 제3기부터 건강면접조사에 HRQoL 관련 설문 항목을 추가하였으며, 도구 타당성이 입증되고 국가간 비교가 가능한 EQ-5D 측정도구를 선택한 바 있다.

3. 건강관련 삶의 질 개념적 모형

노인 만성질환자들의 HRQoL에 영향을 미치는 요인을 설명하기 위한 본 연구의 가설적 모형을 구축하기 위하여, 기존에 연구되어온 삶의 질 모형을 살펴보고자 한다.

3.1 Stuifbergen의 만성질환자를 위한 건강증진 및 삶의 질 모형

Stuifbergen의 만성질환자를 위한 건강증진 및 삶의 질에 대한 개념적 모형(Conceptual model of health promotion and quality of life for people with chronic and disabling conditions)의 초기 개발은 만성질환인 다발성 경화증 환자의 연구로부터 도출되었다. 이 모형은 건강증진 행위와 삶의 질을 설명하기 위하여 세 단계로 구별되었으며, 각 단계에 영향을 주는 요인들이 포함되어 있다(Stuifbergen & Rogers, 1997)(그림 2).

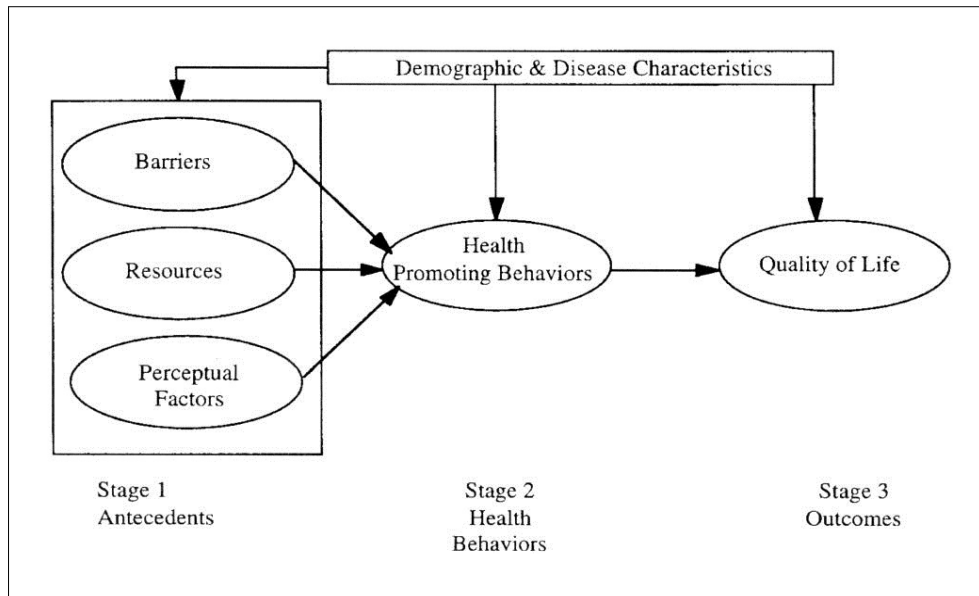


Figure 2. Conceptual model of health promotion and quality of life for people with chronic and disabling conditions

Stuifbergen 모형의 선행요인으로 방해요인, 자원요인, 인지요인이 있다. 방해요인은 개인의 건강증진 행위 선택 시 어려움이나 불편감, 혹은 이용 불가능에 대한 지각으로 건강증진 행위와 삶의 질에 부정적인 영향을 주는 개인의 내적 요인, 대인관계 요인, 환경적 요인을 의미한다(Stuifbergen, 1995). 자원요인은 건강증진 행위의 선택과 사용을 촉진하게 하여 삶의 질에 직간접적으로 영향을 주는 개인적, 환경적 특성을 의미하는 것으로, 건강증진 행위에 직접 영향을 주는 요인으로 사회적 지지 등을 포함한다(Stuifbergen & Rogers, 1997). 인지요인은 건강증진 행위를 시작하려는 가능성에 영향을 주는 특성으로 건강행위에 대한 자기효능감 등을 포함한다(Stuifbergen, 1995).

건강증진 행위는 선행요인과 삶의 질 사이의 중개자와 같은 역할을 하게 된다. 결과적으로 만성질환자들이 재활의 목표로서 변화된 생활양식에의 적응, 적절한 건강유지, 가능한 최상의 기능을 돕기 위해 이성적으로 협동

하는 것을 의미한다(Stuifbergen, 1995).

이 모형의 핵심개념은 삶의 질이다. 이는 삶에 대한 만족도와 안녕, 건강에 대한 개인의 전반적인 느낌에서 유도된 개념이다(Oleson, 1990). 이 모형에서는 선행요인들이 건강증진 행위를 매개로 또는 직접적으로 삶의 질에 영향을 미치게 된다.

Stuifbergen 모형은 초기에 장애를 가진 성인들 간의 건강증진 행위 발생을 설명하기 위하여 Pender (1987)의 건강증진모형에 대한 유용성을 계층적 다중회귀분석(Hierarchical multiple regression analyses) 방법을 통해 검증하였으며(Stuifbergen & Becker, 1994), 이후로 구조방정식분석(Structural Equation Analyses: 이하 SEA) 방법을 통해 다발성경화증 환자(Stuifbergen, 1995)와 소아마비증후군 환자(Stuifbergen, Seraphine, Harrison, & Adachi, 2005)를 대상으로 검증하였다.

3.2 Wilson과 Cleary의 질병과 삶의 질 연계모형

Wilson과 Cleary의 질병과 삶의 질 연계모형(Conceptual model of link between disease and quality of life)은 질병의 원인에 초점을 두는 생의학적 패러다임(biomedical paradigm)과 환자의 기능과 전반적 안녕에 초점을 두는 사회과학적 패러다임(social science paradigm)을 통합하는 건강관련 삶의 질의 개념적 모형이다(I. B. Wilson & P. D. Cleary, 1995)(그림 3). 이 모형은 생리적 요인, 증상, 기능적 건강상태, 전반적 건강에 대한 인식과 삶의 질을 주된 맥락으로 하며, 환자의 개인적, 환경적 특성과의 연결되어 있음을 제시하였다.

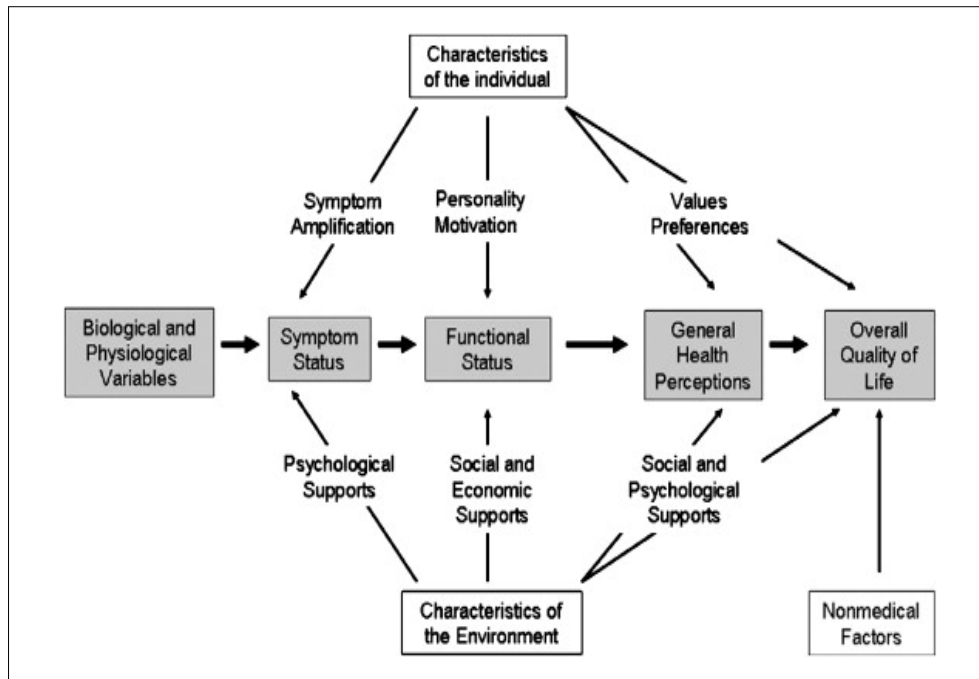


Figure 3. Conceptual model of link between disease and quality of life

개인적 특성은 가족력 같은 생물학적 요소, 성별, 나이, 결혼상태, 종교와 같은 인구학적 요소, 발달요소, 질병에 대한 인지적 평가를 반영하는 역동적인 심리적 요소 등이 있다. 환경적 특성은 사회적 지지와 문화적 특성 같은 사회적 환경과 가정, 이웃, 업무공간 등 건강에 영향을 주는 물리적 환경을 말한다(Ferrans, Zerwic, Wilbur, & Larson, 2005).

이 모형은 암환자(Wettergren, Björkholm, Axdorph, & Langius-Eklöf, 2004), AIDS(Sousa & Kwok, 2006), 관상동맥질환(Höfer et al., 2005), 비만(Ropka, 2012) 등의 다양한 대상에 폭넓게 적용되고 있으며, 질병 특성에 따라 언급한 개념을 전체 또는 부분적으로 이용해서 삶의 질과의 관계를 SEA 방법을 사용하여 검증하였다.

Wilson과 Cleary의 모형은 HRQoL에 미치는 다차원적 영향요인을 이해하게 함으로써 대상자의 돌봄에 합리적인 근거를 제공하였으며, Ferrans 등(2005)은 간호와 보건 분야에서 건강관련 삶의 질 개념 모형의 활용을 위해

문헌고찰을 통해 Wilson과 Cleary의 모형을 수정하여 제시하였다(그림 4).

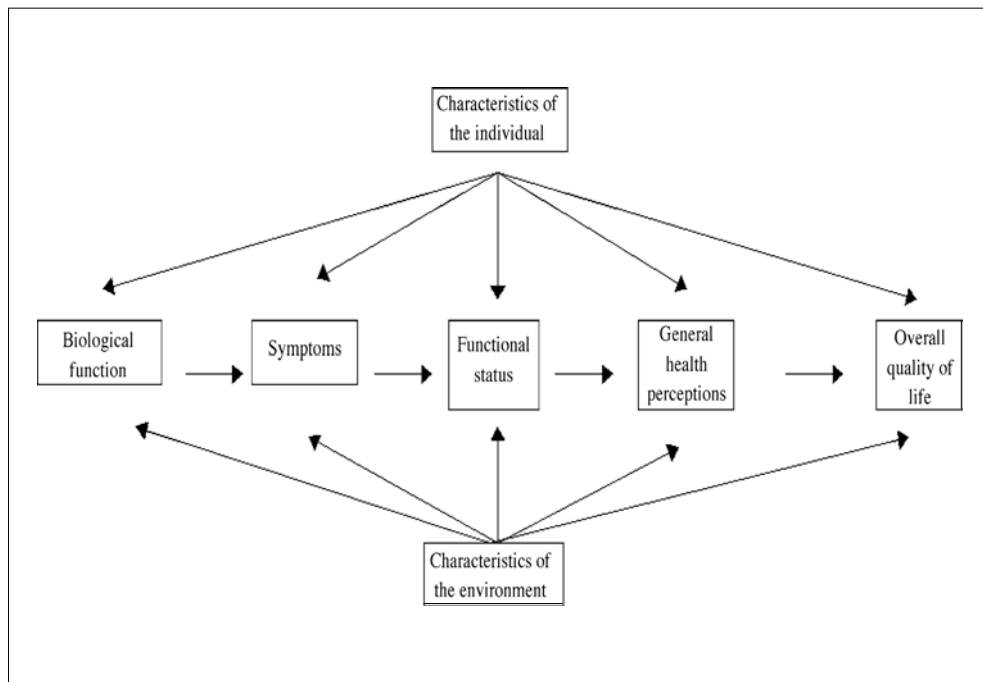


Figure 4. Revised Wilson and Cleary model for health-related quality of life

Wilson과 Cleary의 수정모형은 HRQoL에 영향을 미치는 요인으로 생물학적 요인, 증상, 기능적 상태, 일반적 건강지각으로 보았으며, 개인적 특성과 환경적 특성이 이러한 요인들을 통하여 건강관련 삶의 질에 직간접적으로 영향을 미치는 것으로 설명하였다. 이들은 원래의 Wilson과 Cleary 모형에서 3가지 사항을 변경하였다. 첫째, 생물학적 요인이 개인적 특성과 환경적 특성에 영향을 받기 때문에 화살표를 추가하였다. 둘째, 비의료적 요인들은 개인적 특성 또는 환경적 특성에 포함될 수 있기 때문에 삭제하였다. 셋째, 화살표의 설명 예시는 관계의 특징을 제한할 수 있기 때문에 삭제하였다. 이들은 이러한 수정모형이 연구와 실제 임상에서 유용하며, 사용성

측면에서 모형의 대상자를 의사로부터 간호사와 다른 건강전문가들의 관점으로 확대하였다고 주장하였다.

Daggett 등은 외상성 뇌손상을 입은 참전 용사들을 대상으로 한 HRQoL 관련 연구들을 검토함으로써 현재 지식과의 차이를 밝히고 향후 연구에 대한 제언을 위해 이 수정모형을 적용하였다(Daggett, Bakas, & Habermann, 2009).

4. 노인 만성질환자의 건강관련 삶의 질 영향요인

본 연구는 만성질환을 가진 노인을 대상으로 하지만 HRQoL에 영향을 미치는 요인에 대해서 폭넓게 선행연구를 고찰하였다.

선행연구들은 연구대상을 성인(길선령, et al., 2008; 안홍모, 2010), 노인(김정연, 이석구, & 이성국, 2010; 이동호 & 빈성오, 2011), 여성(김미령, 2006; 손연정, 2007) 등 폭넓게 접근하거나 기러기아빠(차은정, 2012), 기혼 직장여성(손연정, 2007), 취약계층 남성노인(전은영 & 최연희, 2010) 등 특정 대상과 당뇨(Porojan, et al., 2012; 최해림, 2008), 뇌졸중(Alguren, et al., 2012), 뇌성마비(이병희 & 고주연, 2010), 만성폐쇄성폐질환(Peruzza et al., 2003), 만성신부전(Brown, 2012) 등 특정질병으로 한정하여 접근하기도 하였다.

연구방법론에 있어서는 사전에 HRQoL 영향요인들을 선택적으로 제한하여 선정하고 이와 관련된 설문을 통하여 자료를 조사하였으며, 대부분 전통적인 통계방법인 회귀분석을 통한 요인분석(Marchesini et al., 2001), 구조모형 분석(Heckman, 2003; 장선주, 2010; 차은정, 2012) 방법을 사용하였으며, 의사결정 나무(Decision Tree: 이하 DT)(최선영, 2009)의 단일한 방법만을 적용하여 제한적으로 분석한 연구도 있었다.

만성질환을 가진 노인들을 대상으로 한 HRQoL에 관한 연구들을 살펴보면 도시지역 노인의 경우 HRQoL이 신체거동과 통증과 관련된 관절질환,

뇌졸중, 편두통 등의 질환과 높은 관련성을 보였다(남해성, et al., 2000). 만성질환 중 근골격계 질환이 가장 HRQoL이 낮았으며, 주관적 건강상태, 근골격계 질환수, 의학적 건강상태 순으로 영향을 미쳤으며(이동호 & 빈성오, 2011), 근골격계 질환과 복합된 질환의 경우 HRQoL에서 낮은 점수를 보였고, 심리적 상태, 신체적 건강, 소득 순으로 영향력이 높았다(남상권 & 심옥수, 2011). 이처럼 노인 만성질환자들을 대상으로 만성질환의 유형에 국한된 한정적인 범위에서 연구들이 시행되었다.

선행연구들 중 일부에서는 본 연구와 같이 KNHANES 자료를 이용하여 HRQoL(배상열 et al., 2010), 유병율(박혜순 et al., 2003), 영양요인(정지연, 김준수, 최현정, 이가영, & 박태진, 2009), 의료이용(이용철, 임복희, & 박영희, 2010) 등에 관한 연구가 시행되었다. 안소현 (2008)은 노인들의 만성질환과 관련된 식이위험 인자에 관한 연구를 통해 노인의 주요 만성질환인 당뇨, 비만, 고지혈증, 고혈압이 서로 밀접한 연관성을 맺고 있어 합병증 등에 대해 종합적인 관리가 필요하며, 특히 비타민 무기질 급원 식품에 대해 적정량을 섭취할 수 있도록 권장하였다. 서성은(2010)은 비만과 관련성이 높은 만성 퇴행성 질환의 증가로 비만관리 필요성이 증대되어 노인의 건강행위 실천과 비만도와의 관련성 연구를 하였으나 유의한 결과가 도출되지 않았다.

선행연구의 고찰을 통해 HRQoL에 관한 영향요인을 개인적 특성, 환경적 특성, 생리적 요인, 증상경험, 건강증진 행위로 분류하여 정리하면 표 2와 같다.

Table 2. Influencing factors of health related quality of life

Type	Influencing Factors
Characteristics of the Individual	Age, Gender, Race, Education
Characteristics of the Environment	Income, Job, Residence, Marriage, Health Insurance
Physiological Factors	Disease type, Complex disease, Illness days, Sick days
Symptom Experience	Depression, Discomfort, Self-esteem
General Health Perceptions	Perceived Health Status
Health Promoting Behaviors	Smoking, Drinking, Activity, Obesity, Stress, Sleeping time, Walking, Diet

상기 분류별 영향요인에 대해 자세히 살펴보면 다음과 같다.

4.1 개인적 특성과 환경적 특성

국내외적으로 성별, 연령, 인종, 소득 등의 인구학적 요인이 HRQoL과 관련이 있으며, 국내 표본을 대상으로 EQ-5D를 사용한 연구에서도 성, 연령, 교육, 직업종류, 월평균 가계소득, 거주지역에 따라 HRQoL 수준이 다르다고 하였다(Jiang & Hesser, 2006; 성상석 et al., 2004). 구체적으로 고등교육을 받은 대도시 거주 남성 사무직원의 HRQoL이 가장 높았고, 교육수준이 낮고, 육체노동을 하며, 농촌지역에 거주하는 여성의 HRQoL이 가장 낮은 것으로 보고하였다(성상석 et al, 2004).

기존의 선행연구를 살펴보면 소득수준과 교육수준이 높을수록 HRQoL이 높게 나타났고, 남성이 여성보다 HRQoL이 다소 높았다. 또한 경제활동을 하고 있는 경우 HRQoL이 상대적으로 높았으며, 관리자, 전문가, 사무직 등

이 농어업숙련직, 단순노무직, 서비스직보다 HRQoL이 상대적으로 높았다. 반면에 연령이 증가할수록 HRQoL이 감소하였으며, 배우자가 없는 경우, 별거, 이혼, 사별을 경험한 사람일수록 HRQoL이 유의하게 낮았고, 의료급여 대상자인 경우 낮은 EQ-5D index를 나타내었다(신상진, 2010; 정영호, 2011; 한미아 et al., 2008). 박웅섭(2010)은 빈곤층 노인 수급자의 HRQoL에 영향을 미치고 있는 소득, 연령, 스트레스 수준, 만성질환 여부, 치료 경험 부재 등의 노인 수급자 특성에 맞는 사회복지 서비스의 연계 및 발굴이 필요하다고 하였다.

전기노인(60대), 중기노인(70대), 후기노인(80대 이상)에서 모두 배우자가 있을수록 자녀수가 많을수록 HRQoL이 높았으며, 자녀수와 동거유형, 만성질환수에 유의미한 차이가 있었고, 비동거 자녀와의 간접접촉과 친구나 친척 및 이웃과의 접촉빈도에서 유의미한 차이가 있다고 보고하였다(최윤진, 2009). 또한 노인들에 있어서 사회적 지원과 사회활동의 참여가 HRQoL에 영향을 주는 것으로 대다수의 선행연구에서 제시하고 있다(조은주, 2009). 농촌 노인의 사회적 지지와 HRQoL 연구에서 동거 가족수와 대상자의 사회적 지지정도와 HRQoL과의 관계가 유의하게 나타났다(강경숙, 2003).

노인의 HRQoL에 영향을 미치는 일반적 특성은 연령, 성별, 배우자유무, 가족관계, 종교, 교육수준, 사회활동, 경제상태 등의 요인들로 다시 세분할 수 있으며, 만성질환 노인의 경우는 성별, 교육 정도, 배우자 동거유무, 주간호 제공자가 HRQoL의 영향요인으로 보고되었다(남상권 & 심옥수, 2011).

만성질환이 우울증세에 미치는 위험효과에 대한 사회관계망의 효과를 검증한 결과, 가족관계는 만성질환 유무와 관계없이 그 자체로 노인의 우울증세를 감소시키는 주효과가 있었으며, 사회단체활동 중에는 친목모임과 자원봉사활동이 만성질환으로 인한 우울증세를 완충하는 효과가 있는 것으로 나타났다. 이로부터 일상생활이 불편한 정도의 만성질환 유병 노인은 신체질환과 우울증세를 통합적으로 접근하는 의료적 개입을 중심으로, 건강한

노인이나 일상생활에 불편이 없는 정도의 노인에 대해서는 사회적 통합을 강화하는 사회교육과 복지제도 활성화 등 사회적 개입이 유효하다고 제안하였다(이미숙, 2010).

4.2 생리적 요인

노인에서 HRQoL은 가족과 친구들을 포함한 사회적 지지나 개인 삶의 가치관을 높여주는 종교생활 그리고 경제수준 등의 여러 영향요인이 있지만 그 중에 가장 큰 영향을 주는 요인은 건강문제다(Larson, 1978). 신체적 건강은 노인의 HRQoL에 긍정적인 영향을 미치는 유의미한 변인으로 나타나고 있다. 질병이 없는 노인은 질병이 있는 노인보다 HRQoL이 상대적으로 높으며, 노년기의 건강은 HRQoL에 영향을 주는 중요한 요소라고 하였다(Bowling et al., 2003).

정영호(2011)는 만성질환 중에서 천식이 HRQoL이 0.850으로 가장 낮았으며, 관절염 0.854, 디스크 0.879, 당뇨병 0.880, 고혈압 0.888로 HRQoL이 낮은 것으로 나타났다고 보고하였다. 길선령(2008)은 HRQoL 수준과 개별질환과의 연관성을 KNHANES를 통해 분석하였으며, 그 결과 뇌졸중군의 EQ-5D index 감소가 가장 컸다고 보고하였다. 신상진(2010)은 EQ-5D index가 0 미만으로 응답한 경우 폐암, 뇌졸중, 대장암, 폐결핵, 천식이 높은 빈도를 나타내었으며, 질환의 이환일수와 외병일수가 전체적으로 높게 나타났다고 보고하였다.

선행연구에서 노인의 HRQoL에 영향을 미치는 요인으로서 만성질환, 만성질환에 의한 일상생활의 어려움 등을 제시하고 있다(김미령, 2006). 해당 연구에서는 만성질환 노인의 경우는 HRQoL의 영향요인으로 주간호 제공자와 동반 질환수가 중요한 특성으로 조사되었으며, 만성질환 유형과 만성질환수에 따라 HRQoL에 미치는 영향이 다르다고 보고하였다. 만성질환 노인의 HRQoL은 순환기계, 순환기계-내분비계, 내분비계, 근골격계, 순환기계-근골격계, 내분비-근골격계, 순환기-내분비-근골격계 질환 순으로 근

골격계 질환이 포함되는 복합적인 질환의 수가 증가할수록 HRQoL 점수가 낮아지는 것으로 보고하였다(남상권 & 심옥수, 2011).

만성질환자의 HRQoL에 영향을 미치는 요인에 관한 기존의 연구로서, Rubin & Peyrot (1999)는 당뇨 질환자는 만성질환자가 아닌 개인에 비해 HRQoL이 현저히 낮고, 합병증 여부가 HRQoL에 영향을 미치는 가장 중요한 변수라고 제시하면서, 임상적, 교육적 중재가 환자의 건강상태 증진과 자신의 조절력 인식을 향상시킴으로써 HRQoL을 증진시킬 수 있다고 보고하였다. 또한 Burckhardt (1985)는 관절염 환자를 대상으로 HRQoL을 연구한 결과 자아 존중감, 내적 통제위, 질병에 대한 부정적인 태도, 사회적 지지 등의 변수가 HRQoL을 설명하는 유의한 변수로 보고하였다.

4.3 증상경험

노년기에 나타날 수 있는 심리적 변화로 우울요인을 들 수 있는데, 이는 노년기에 신체활동과 건강기능의 쇠퇴, 은퇴로 인한 무력감 등으로 인하여 우울을 경험할 수 있기 때문이다.

우울과 관련된 선행연구들을 살펴보면 농촌 노인의 HRQoL에 우울이 직접적인 영향을 미치며(손신영, 2006), 전기노인과 후기노인의 경우에도 우울은 HRQoL이 직접적인 영향이 있는 것으로 나타났다(최윤진, 2009). 또한 일상생활 활동이 어려운 노인의 경우 우울을 많이 느낀다고 하였고, 사회적 지지가 낮고, 자아 통제감이 낮으며, 주관적 건강상태가 나쁘다고 생각한 노인의 우울이 높다고 제시하였다(박상규, 2006). 대도시 지역 노인들을 대상으로 한 연구에서 연구자들은 우울과 HRQoL 간에는 역상관 관계가 있다고 하였다(김태현, 김동배, 김애순, 김미혜, & 이영진, 1999).

질병으로 인해 신체적 불편감이 만성화되고 기능장애가 있는 경우, 전반적인 건강의 변화로 인하여 여가활동의 참여가 제한되고 신체적, 정서적 불편감을 초래하고 자기 수용과 사회적 적응의 어려움으로 자존감이 낮아지며 HRQoL에 부정적인 영향을 받게 되므로 신체적 치료뿐만 아니라 이들의

HRQoL을 높이도록 돕는 것이 필요하다(Holmes, 1985).

불편감이나 건강상태 또는 질병 유무가 HRQoL과 상관관계가 있다는 선행연구가 있었으며, 류마치스성 관절염 환자들의 불편감 정도와 HRQoL간에 음의 상관관계를 보였던 연구결과와는 차이가 있었으나, 말기 신부전 환자, 관절염 환자들의 HRQoL이 정상보다 저하되지 않았음을 보고한 결과도 있었다(Laborde & Powers, 1985). 이와 같이 질병상태에서 HRQoL이 저하되지 않은 이유는 만성질환자의 경우 질병의 초기단계를 벗어나게 되면 습관화된 불편감에 적응하거나 질병에 대한 대응능력이 강화되기 때문이라고 볼 수 있다(김명자 & 송경애, 1991).

4.4 일반적 건강 지각

주관적 건강 또는 지각된 건강(Subjective Health, Perceived Health)은 개인이 스스로 평가하는 자신의 건강(Self-rated Health, Self-reported Health) 수준을 말한다. 주관적 건강은 의학적, 보건학적 건강수준의 유의미한 예측요인이며, 생물학적, 정신적, 사회적, 기능적 건강수준을 포괄하는 전반적 건강을 측정하는 타당한 척도로 받아들여진다(Ocampo, 2010; 신경림 & 김정선, 2004).

정신적 스트레스가 있는 경우, 주관적 건강인식이 나쁠수록 HRQoL이 유의하게 낮은 것으로 나타났으며, 스트레스를 많이 인지하는 사람일수록 본인의 건강상태를 완전한 건강으로 응답할 확률이 낮았으며, 낮은 EQ-5D index를 나타내었다(한미아, 2008; 신상진, 2010).

선행연구들은 노인 스스로 지각한 건강상태가 좋을수록 HRQoL이 높게 나타났음을 제시하였다. 주관적 건강은 객관적인 영향상태에 영향을 주며 나아가 생활만족도와도 연관되어 있고, 노인의 자아존중감에도 영향을 주는 것으로 나타나고 있다(최수정, 2001).

4.5 건강증진행위

건강한 생활습관과 예방적 건강행위는 성공적 노화를 통한 삶의 질 향상에 있어 중요한 요소이다(차남현, 2010). 노인의 건강행위와 HRQoL 간에는 유의한 상관관계가 있는 것으로 나타났으며, 비만, 음주, 신체활동, 스트레스와 같은 건강위험 행태요인이 EQ-5D index에 영향을 미친다고 하였다(신상진, 2010; 최정숙, 권성옥, & 백희영, 2004).

현재 흡연자들이 과거 흡연자와 비흡연자에 비해 HRQoL이 낮았으며(Jiang & Hesser, 2006), 정상체중에 비해 고도비만일 경우 HRQoL이 감소하였고(정영호, 2011), 비만인 사람이 체중을 감소하였을 때 HRQoL이 좋아지는 것으로 나타났다(Y. T. Lim, Park, Kim, Shin, & Kim, 2001). 홍인옥(2011)은 흡연, 음주, 운동, 비만 여부에 따른 EQ-5D index를 비교 분석한 결과 여성의 경우 현재 흡연자이면서 비만할수록 유의하게 HRQoL이 낮게 나타났으며, 운동을 불규칙적으로 하는 경우(남성 94%, 여성 95%) 규칙적인 운동을 하는 경우에 비해 HRQoL이 낮은 것으로 보고하였다.

관절염 여성의 경우 건강증진이 HRQoL에 직접적인 효과를 가지고 있었고, 질병에 대한 심각성의 정도와 자기효능감이 건강증진을 통해 HRQoL에 간접적인 영향을 미치는 것으로 나타났다(Oh, 1993). 당뇨병 노인의 HRQoL을 설명하는 주요 변수는 한번에 마시는 음주량, 스트레스 인지율, 자살 생각률, 걷기 지속시간이었으며, 당뇨병이라는 만성질환과 노년기라는 이중의 고통을 겪고 있는 당뇨병을 가진 노인 스스로 자신의 생활 습관에 대한 관리를 효율적으로 할 수 있는 교육과 간호중재의 필요성을 제시하였다(신지원, 박용경, 서순림, & 김지은, 2011).

5. 보건의료 분야에서의 기계학습 적용

기계학습이란 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 말하며, 환경과의 상호작용에 기반한 경험적인 데이터로부터 스스로 성능을 향상시키는 시스템을 연구하는 과학과 기술로 정의된다(장병탁, 2007).

기계학습이라는 용어는 Samuel (1959)이 게임보드의 패턴으로 정의된 특징과 가중치를 변경함으로써 게임을 학습하는 방법을 제안하면서 처음 문헌에 등장하였다. 이후 계산이론과 접목된 기계학습이 이론적인 틀을 갖추면서 새로운 연구분야로 자리잡기 시작한 것은 1980년대 중반에 이르러서이다. 1990년대 중반 이후에는 인터넷과 웹이 활성화되고, 데이터마이닝이 새로운 산업으로 등장하면서 기계학습은 거의 모든 분야에서 핵심기술로 자리매김하게 되었다(장병탁, 2007).

기계학습은 문자 인식(김수형, 2007), 문서 분류(Sebastiani, 2002), 정보 검색(Y. Yu & Wang, 2002), 이미지 분석(L. Zhang, Lin, & Zhang, 2001), 생체정보 분석(김광준 & 이형근, 2011), 생물정보 분석(Shipp et al., 2002), 임상진단과 예측(Chan et al., 2002), 금융자료 분석(박강희 & 신현정, 2010) 등 다양한 분야에 응용되고 있다. 그 응용 분야와 적용 사례를 살펴보면 다음과 같다(Allison, Cui, Page, & Sabripour, 2006; B.-T. Zhang, 2010) (표 3).

Table 3. Applications and case of machine learning

Applications	Case
Internet information retrieval	Text mining, Web log analysis, Spam filter, Document classification, Filtering, Extraction, Summary, Recommendation
Computer vision	Character recognition, Pattern recognition, Object recognition, Face recognition, Scene change detection, Image restoration
Lexical processing	Voice recognition, Word ambiguity removing, Translation word selection, Grammar learning, Dialogue pattern analysis
Mobile Human computer interaction	Motion recognition, Gesture recognition, Sensor information recognition, Compensating tremble
Bioinformatics	Gene recognition, Protein classification, Gene regulatory networks analysis, DNA chip analysis, Disease diagnosis
Biometrics	Iris recognition, Heart rate measurement, Blood pressure measurement, Blood glucose measurement, Fingerprint recognition
Computer graphics	Animation based data, Character motion control, Inverse kinematics, Behavior evolution, Virtual reality
Robotics	Obstacle recognition, Object classification, Map building, Autonomous driving, Path planning, Motor control
Services	Customer analysis, Market cluster analysis, Customer management, Marketing, Product recommendation
Manufacturing	Anomaly detection, Energy consumption prediction, Scheduling analysis, Error analysis and classification

기계학습은 이를 활용하여 얻고자 하는 정보의 종류에 따라 주어진 데이터의 특성 혹은 패턴의 분석을 목표로 하는 무감독 학습(Unsupervised Learning)과 주어진 데이터를 효과적으로 분류하는 함수의 도출을 목표로

하는 감독 학습(Supervised Learning)으로 크게 분류할 수 있다.

의료기술 및 환경의 발전으로 의료분야에서도 다양한 대용량 자료들이 생겨나고 있으며, 이러한 자료를 어떻게 분석해서 활용할 것인가에 대한 해결방안을 모색하고 있는 것이 사실이다. 의료 전문가의 해석과 판단도 중요하지만 사람이 통제하기 어려운 데이터를 분석하고 그 속에서 의미있는 정보를 추출하기 위해서는 전문가의 지식과 기계적인 접근방법을 상호보완적으로 적용한다면 보다 효율적인 성과를 얻어낼 수 있을 것이다.

최근 ML을 이용한 새로운 예측방법들이 다양한 임상현장에서 적용되고 있다. 그러나 성능평가에 있어서는 결측 및 이상 자료, 다른 형태의 자료처리, 차원감소 기법 등의 자료처리에 따라 결과가 달라질 수 있어 아직도 논의 중이다. 기존의 방법과 ML 기법은 서로 상호보완적으로 사용될 수 있으며, 전통적인 통계방법인 회귀분석을 이용하여 ML 자료의 차원축소를 실시한 연구도 있다(Kim, Kim, & Park, 2011).

본 연구에서는 최적의 HRQoL 예측모델을 찾기 위한 다양한 시도로서 전통적인 LR 방법과 새로운 전산적 기법인 ML 기법 중에 분류 및 예측에 적합한 DT, RF, SVM의 3가지 방법을 함께 적용하여 예측모델의 성능을 비교하였다. 어떤 방법이 가장 우수하다고 판단하기는 어려우며, 데이터의 특성 및 적용 목적 등에 따라 상황에 맞는 최적의 모델을 개발할 필요가 있다. 3가지 방법에 대하여 각각의 정의와 보건 의료 분야에서의 선행연구들을 살펴보면 다음과 같다.

5.1 의사결정나무(Decision Tree)

DT 기법은 순환적 분할(Recursive partitioning) 방식을 이용하여 나무를 구축하는 기법으로 가장 상단에 위치하는 뿌리마디(root node), 속성의 분리기준을 포함하는 내부마디(internal nodes), 마디(node)와 마디를 이어주는 가지(link), 그리고 최종 분류를 의미하는 잎(leaf)으로 구성되며, 분류(Classification)와 예측(Prediction)에 주로 사용된다.

DT 기법은 노이즈에 강하고, 논리합 표현을 학습하는 이산 함수를 유도하는 방법이며, 이산값으로 표현되는 데이터로부터 규칙을 학습하는데 적합하다고 할 수 있다(Mitchell, 1997). 이 방법은 분류 또는 예측의 과정이 나무구조에 의한 추론규칙에 의해서 표현되기 때문에 다른 방법들에 비해서 연구자가 그 과정을 쉽게 이해하고 설명할 수 있다는 장점을 갖는다(이현주, 2011). 반면, 분류모델을 생성하는데 있어서 과대적합(over-fitting)이나 과소적합(under-fitting) 문제가 발생하여 결과적으로 분류기법의 예측력도 감소할 수 있다(서동혁 et al., 2012). 또한 모델을 구축하는데 사용되는 표본의 크기에 매우 민감하여 보다 정확한 모델을 구축하기 위해서는 서로 상이한 값을 갖는 레코드들을 가능한 한 많이 포함하는 데이터가 필요하다(장남식, 홍성완, & 장재호, 2000).

DT의 보건의료 분야 적용연구들을 살펴보면 다음과 같다. 신영희(2009)는 DT 분석방법을 이용하여 도시 노인의 성공적인 노화 관련 요인을 규명하였는데 성별, 학력, 배우자, 운동, 우울, 만성질환 등이 이에 해당되었다. 최동운(2010)은 DT 알고리즘을 이용한 생체 데이터마이닝 모듈을 개발하였으며, 이를 통해 건강상태가 비만, 비만징후, 정상, 응급상황 여부를 판단하며, 비만으로 인하여 발생할 수 있는 만성질환 발생을 예방하여 건강한 삶을 유지할 수 있도록 도울 수 있다고 주장하였다. 이외에도 DT 방법을 적용한 심전도 데이터 분류(신동규, 2012), 심장질환의 위험수준 예측을 위한 임상 의사결정지원시스템(Anooj, 2011), 파킨슨 질병의 치료 가이드라인 관리를 위한 알고리즘 개발(Olanow, Watts, & Koller, 2001) 등 다양한 연구들이 진행되었다.

5.2 랜덤 포레스트(Random Forest)

ML 분류기법의 하나인 RF는 Breiman에 의해 개발된 것으로 트리(tree) 타입의 분류기(classifier)들을 이용한 앙상블(ensemble) 분류기법이다. 앙상블 분류기법은 경험적인 객체들로 몇 개의 분류기를 훈련시키고, 그 분류

기들의 결정을 결합하여 가중 투표를 통해 비경험적인 객체를 분류하는 방법이다. 앙상블 기법은 트리들의 결과를 합산하여 더욱 신뢰도 높고 안정적인 분류결과를 내놓는다는 장점이 있다(Duda, 2002; Lepetit & Fua, 2006). 이 방법은 단일 나무를 만드는 대신 여러 개의 나무로 확장시켜 분류의 정확도를 높인다(그림 5).

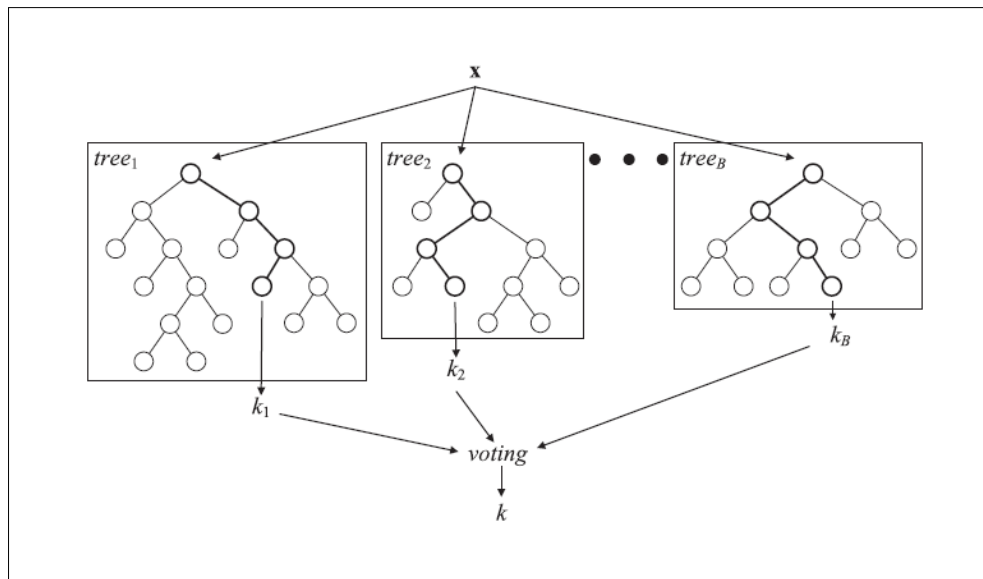


Figure 5. A general architecture of a random forest (Verikas, Gelzinis, & Bacauskiene, 2011)

RF는 표본의 수가 적고 모형에 적용되는 독립변수의 수가 많을 때, 독립 변수와 특정질환 간의 연관성을 찾는데 예측력이 탁월한 동시에 과적합을 방지할 수 있는 모형으로 알려져 있다(Lunetta, Hayward, Segal, & Van Eerdewegh, 2004; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). 이는 대수의 법칙에 의해 나무숲(forests)이 커질수록, 즉 나무의 수가 많아질수록 일반화 오류가 특정값으로 수렴한다는 데서 온 결과이다(한은정, 2005).

RF는 크고 복잡한 시스템에서도 중요한 데이터를 선택하여 분류함으로써 빠르고 정확한 평가를 한다는 점에서 대용량 데이터를 다루는 바이오 정보

학 분야에서 최근 많이 적용되고 있다. 의료분야의 의사결정에 RF 기법을 적용한 국내외 연구로는 이미지와 유전정보를 결합하여 알츠하이머 질환을 분류한 연구(Gray, Aljabar, Heckemann, Hammers, & Rueckert, 2012), 만성폐쇄성폐질환 환자의 건강상태 분류연구(Bellos, Papadopoulos, Rosso, & Fotiadis, 2012), 심혈관 질환의 사망률 예측연구(Austin, Lee, Steyerberg, & Tu, 2012), 단백질 RNA 결합 예측연구(Xin & Xiao, 2012) 등에서 효과적인 결과를 보여주었다.

5.3 지지벡터기계(Support Vector Machine)

SVM은 두 개의 범주를 구분하는 문제를 해결하기 위해 고안된 방법으로 두 개의 클래스의 구성 데이터들을 가장 잘 분리할 수 있는 결정면(Decision surface)을 찾는 모델이다(Ward, McGuffin, Buxton, & Jones, 2003; W. Yu, Liu, Valdez, Gwinn, & Khoury, 2010)(그림 6). 이는 1970년대 후반에 Vladimir Vapnik에 의해 제안된 후 많이 주목 받지 못하다가 1990년대 들어 다양한 분야에서 신경망보다 분류율이 좋은 것으로 평가되고 있다.

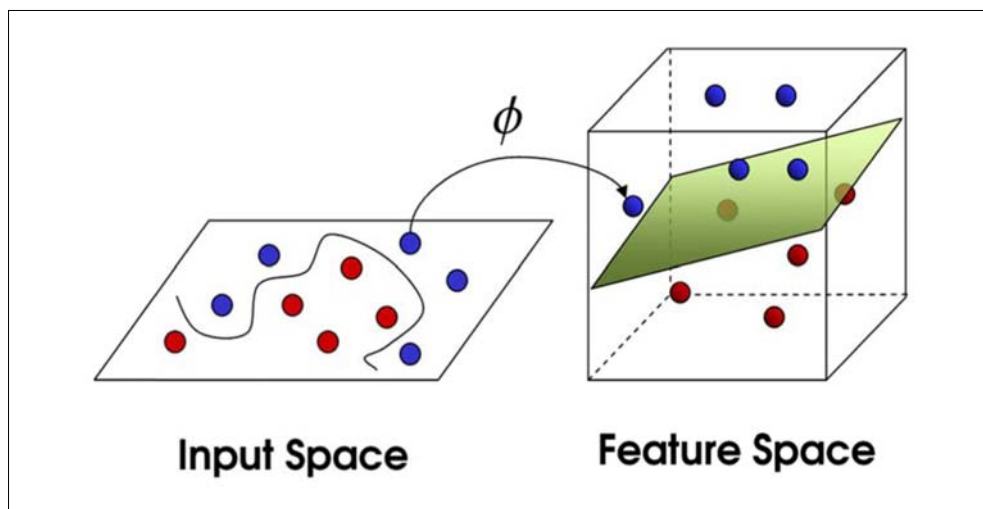


Figure 6. Pattern classification by an SVM(Takahashi, 2012)

주어진 데이터를 분류하기 위한 분류기를 구축하는 방법론은 베이지안 분류기(Bayesian Classifier), 인공신경망(Artificial Neural Network: 이하 ANN), SVM 등 다양한 방법론이 존재한다. 타 방법론의 경우는 주어진 데이터들을 정확하게 분류하는 분류 함수 혹은 분류기를 찾아내면 분류기를 생성하는 작업을 멈추고, 현재까지 찾아진 분류기를 결과로 제공하는 반면 SVM은 주어진 데이터를 분류하는데 만족하지 않고, 각기 다른 부류에 속하는 데이터들의 거리를 최대한 멀리하는(Maximal margin) 분류기를 구축함을 그 목표로 한다.

이로 인해 다른 분류 방법론과 비교하여 SVM의 성능이 최소한 같거나 더 우월하다는 연구결과가 제시되고 있다(Chen et al., 2007; 김상윤, 2011; 송미혜, 이진, 조성필, & 이경중, 2005). 오늘날 SVM은 패턴분석을 위한 강력한 자동화 알고리즘으로써 기계학습 분야에서 주목을 받고 있으며, 다양한 관련 연구들이 문자 및 음성 인식(Solomonoff, Quillen, & Campbell, 2004), 문서 검색(Cao et al., 2006), 이미지 인식(Cusano, Ciocca, & Schettini, 2003), 의학 진단(Polat, Güneş, & Arslan, 2008) 등에 관하여 진행되고 있다.

SVM은 데이터를 분리하는 초평면(Hyperplane) 중에서 데이터들과 거리가 가장 먼 초평면을 선택함으로써 훈련 데이터의 과적합 문제를 방지할 수 있다. 그렇기 때문에 의료 데이터와 같이 표본의 수가 변수의 수보다 작은 대용량 데이터를 분석하는데 적합하다(서동혁, et al., 2012).

보건의료 분야에서의 SVM 적용연구를 살펴보면 다음과 같다. 유재학(2009)은 SVM을 이용하여 신노년층의 생활 만족도에 영향을 미치는 결정적인 19개의 주요 설문 문항들(가족 지지, 수행능력, 욕구충족, 생산적 활동 등)을 도출함으로써, 핵심 문항으로만 구성된 보다 정확한 예측모형을 구축하였다.

김상윤(2011)은 전립선 경직장 초음파 유도 하 조직검사의 항목들을 이용하여 전립선암의 수술 전 병기결정을 예측할 수 있는 SVM 모델과 ANN 모델을 만들고, 이 모델들의 정확도를 비교한 결과 SVM 모델이 임상결정

모델로 더 우수하다고 판단하였다.

이외에도 SVM을 적용하여 Single Nucleotide Polymorphism (SNP) 정보 분석을 통해 만성간염의 초기진단 예측(김동희, 함기백, & 김진, 2006), 심전도 분석을 통해 부정맥 신호분류(이금분, 2010), SPECT 이미지 분류를 통한 알츠하이머병의 진단(Fung & Stoeckel, 2007), 심부전 환자의 약물복용 이행예측(Son, Kim, Kim, Choi, & Lee, 2010)에 효과적이었던 연구들이 보고되고 있다.

Ⅲ. 개념적 기틀 및 연구가설

1. 개념적 기틀

본 연구의 개념적 기틀은 Stuijbergen의 만성질환자를 위한 건강증진 및 삶의 질 모형(Conceptual model of health promotion and quality of life for people with chronic and disabling conditions)과 Wilson & Cleary의 질병과 건강관련 삶의 질 수정모형(Revised Wilson and Cleary model for health-related quality of life)을 근간으로 하고 있다.

Stuijbergen 모형은 삶의 질, 만성질환, 건강증진의 개념이 포함되어 있으며, Wilson과 Cleary 수정모형은 생물학적 기능, 증상, 기능적 상태, 전반적 건강에 대한 인식과 삶의 질을 주된 맥락으로 하며, 개인적, 환경적 특성과의 연결되어 있음을 제시하였다.

본 연구의 주요 개념으로서 Stuijbergen 모형에서의 선행요인인 방해요인, 자원요인, 인지요인에 대한 내용이 본 연구의 대상인 KNHANES 자료에는 직접적으로 측정되지 않았기에 제외하였으며, 인구사회학적 특성과 질병 특성, 건강증진 행위만을 포함하였다. Wilson과 Cleary 수정모형에서도 본 연구대상에서 측정되지 않은 기능적 상태에 관한 요인은 제외하였으며, 생물학적 기능, 증상, 일반적 건강지각, 개인적 특성과 환경적 특성을 포함하였다. Stuijbergen 모형에서 제시하는 인구사회학적인 특성과 질병 특성은 Wilson과 Cleary 수정모형의 개인적 특성, 환경적 특성, 생물학적 기능, 증상으로 나누었으며, 정확한 의미 전달을 위하여 생물학적 기능은 생리적 요인으로, 증상은 증상경험으로 바꾸어 표현하였다. Stuijbergen 모형에서 제시된 건강증진 행위는 선행요인과 삶의 질 사이의 중개자와 같은 역할을 하는 요인으로 판단되기 때문에, Wilson과 Cleary 수정모형에서 삶의 질 전 단계 요인으로 가설적 모형에 포함하였다.

본 연구의 개념적 기틀을 구축하기 위하여 아래의 그림과 같이 Stuijbergen

과 Wilson과 Cleary 수정모형에서 본 연구모형에 포함된 개념(실선 표시)과 포함하지 않는 개념(점선 표시)을 구분하였다(그림 7).

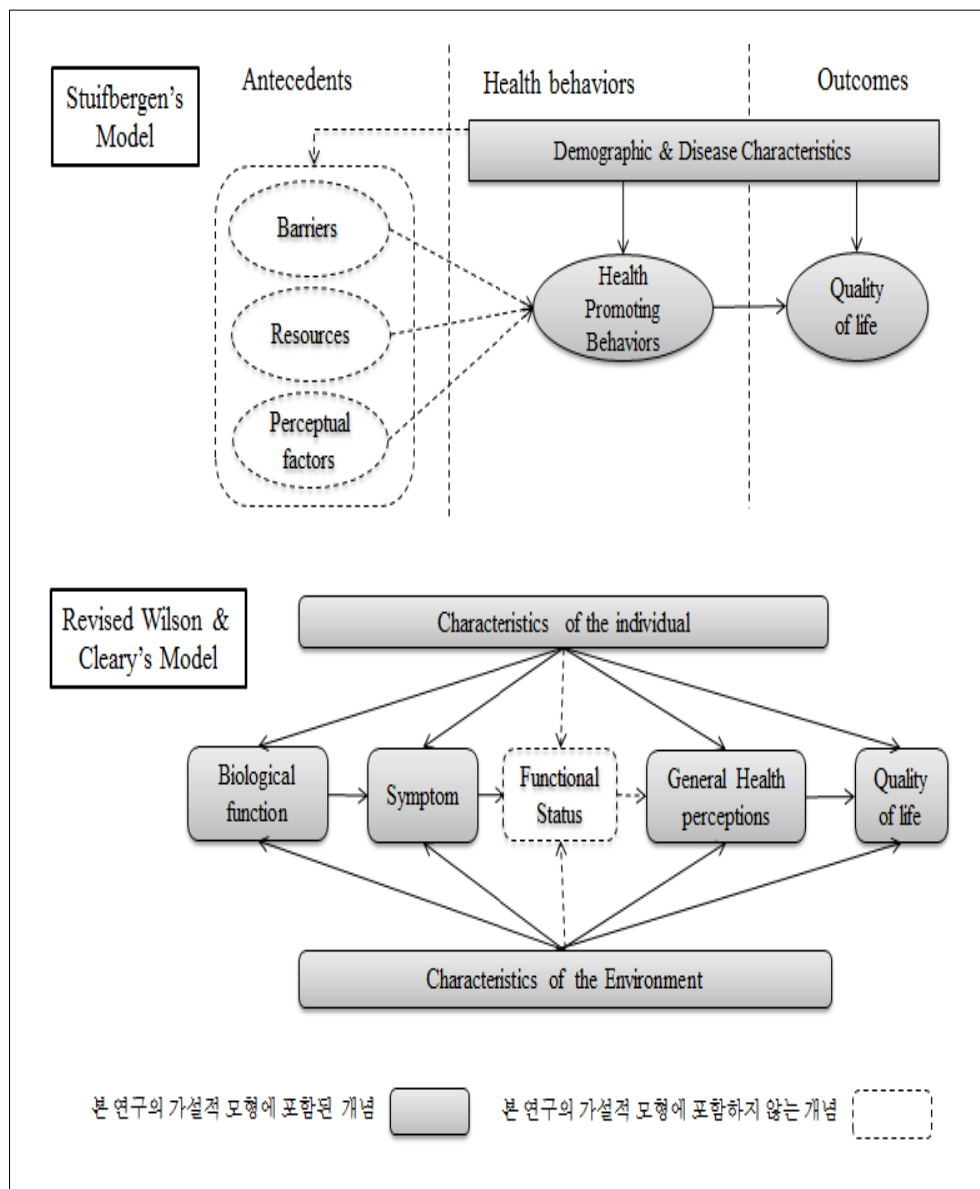


Figure 7. Construction of hypothetical model

본 연구의 주요 개념들은 노인 만성질환자들의 HRQoL 예측모형을 구축하기 위하여 문헌고찰에 근거하여 개인적 특성과 환경적 특성, 생리적 요인, 증상경험, 일반적 건강지각, 건강증진행위, 그리고 삶의 질이라고 가정하였다. 이를 도식화한 연구의 개념적 기틀은 다음과 같다(그림 8).

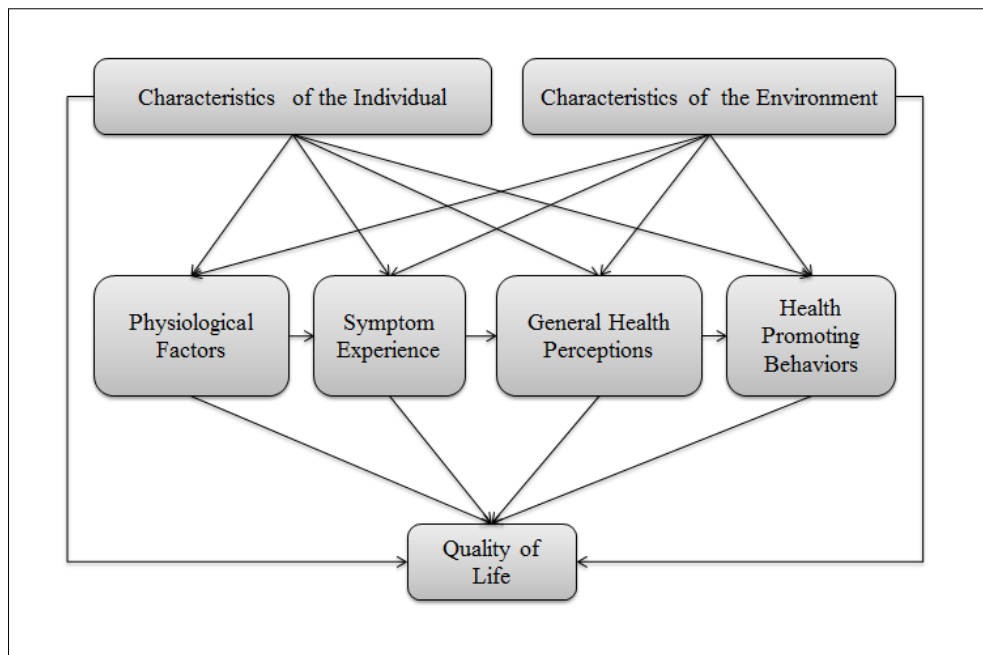


Figure 8. Conceptual framework of prediction model of health related quality of life in elderly with chronic diseases

2. 가설적 모형

본 연구는 위에서 제시한 Stuijbergen 모형과 Wilson과 Cleary 수정모형을 합성하여 구축한 개념적 기틀과 문헌고찰을 이용하여 다음과 같은 가설적 연구모형(그림 9)을 구성하였다.

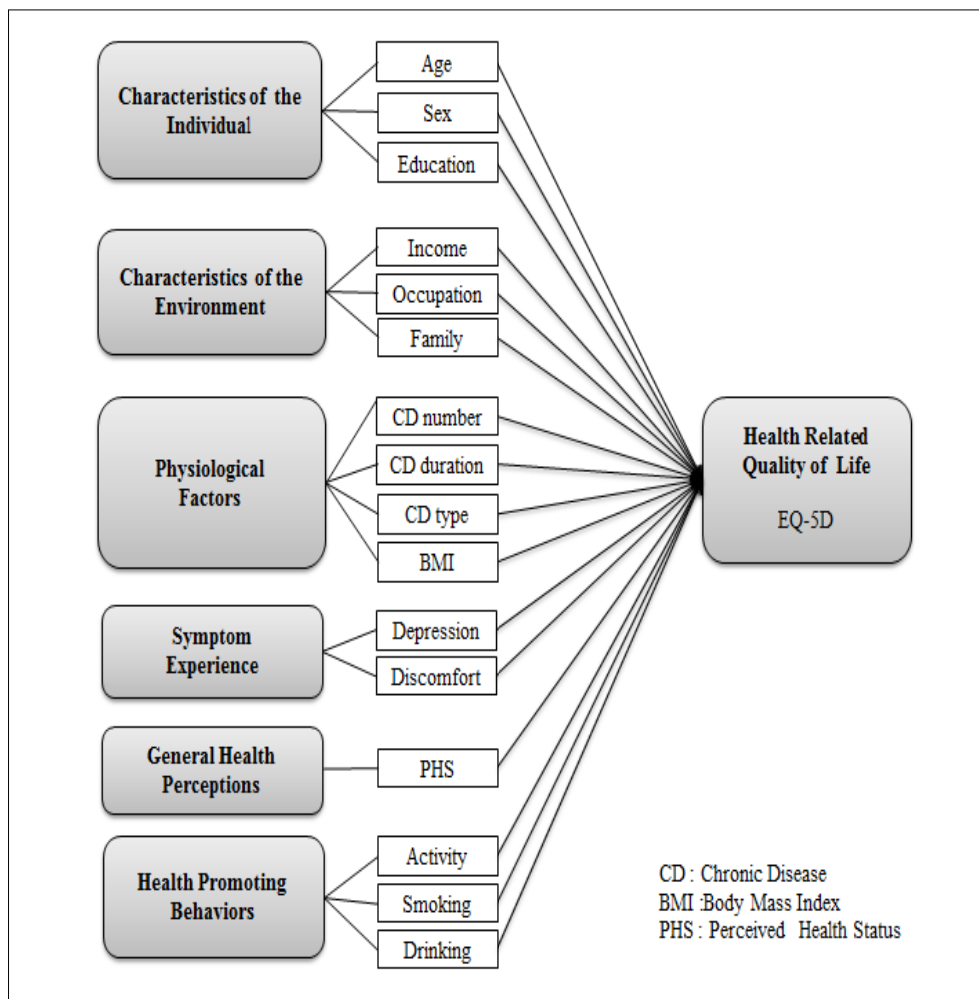


Figure 9. Hypothetical model of health related quality of life in elderly with chronic diseases

가설적 모형을 구축하기 위해 선행연구에 근거하여 HRQoL에 직접적으로 영향을 미치는 요인들을 선택하였으며, 그 요인들은 다음과 같다. 가설적 모형에서 제시된 변수는 연령, 성별, 학력, 소득, 직업, 가족, 만성질환 개수, 만성질환 기간, 만성질환 종류, 체질량지수, 우울, 불편감, 주관적인 건강상태, 활동량, 흡연, 음주로 총 16개이다.

첫째, 개인적 특성인 연령, 성별, 학력은 HRQoL에 영향을 줄 수 있다(Jiang & Hesser, 2006; 정상석, et al., 2004; 정영호, 2011).

둘째, 환경적 특성인 소득, 직업, 가족은 HRQoL에 영향을 줄 수 있다(한미아, 2008; 박웅섭, 2010; 신상진, 2010; 최윤진, 2009; 남상권, 2011; 이미숙, 2010).

셋째, 생리적 요인인 만성질환 개수, 만성질환 기간, 만성질환 종류, 체질량지수(Body Mass Index: 이하 BMI)는 HRQoL에 영향을 줄 수 있다(Bowling et al., 2003; 정영호, 2011; 길선령, 2008; 신상진, 2010; 최윤진, 2009; 김미령, 2006; 남상권, 2011).

넷째, 증상경험인 우울과 불편감은 HRQoL에 영향을 줄 수 있다(Holmes, 1985; Laborde & Powers, 1985; 손신영, 2006; 최윤진, 2008; 박상규, 2006; 김정연, 2010).

다섯째, 일반적 건강지각인 주관적 건강상태는 HRQoL에 영향을 줄 수 있다(Ocampo, 2010; 신경림, 2004; 한미아, 2008; 신상진, 2010; 최수정, 2001).

여섯째, 건강증진 행위인 활동량, 흡연, 음주는 HRQoL에 영향을 줄 수 있다(Jiang & Hesser, 2006; 최정숙, 2004; 홍인옥, 2001; 신지원, 2011).

3. 연구가설

가설적 모형을 토대로 다음과 같은 가설을 설정하였다.

1. 연령은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
2. 성별은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
3. 학력은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
4. 소득은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
5. 직업은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
6. 가족은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
7. 만성질환 개수는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
8. 만성질환 기간은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
9. 만성질환 종류는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
10. 체질량지수는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
11. 우울은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
12. 불편감은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
13. 주관적 건강상태는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
14. 활동량은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
15. 흡연은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.
16. 음주는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.

IV. 연구 방법

1. 연구 단계

본 연구는 다음의 단계에 따라 진행하였다(그림 10).

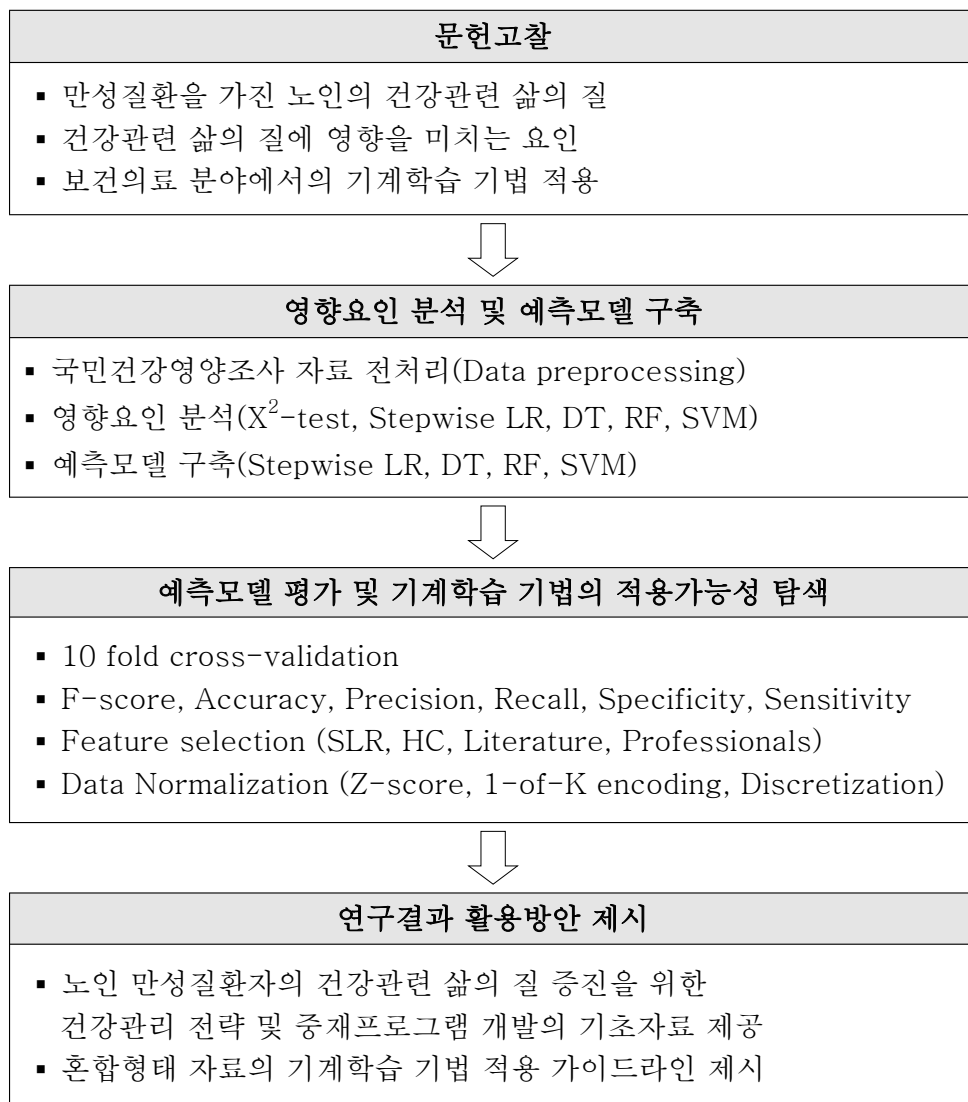


Figure 10. Research process

2. 연구 대상

본 연구에서는 2008~2010년에 실시된 제 4기 및 5기 KNHANES 자료를 이용하였으며, 이는 질병관리본부에서 주관하는 건강설문조사, 검진조사 및 영양조사로 구성된 전국 규모의 연중 조사(전국에서 임의로 약 1만명을 추출)를 통해 얻어진 결과이다. KNHANES는 국민건강증진법에 따라 국민건강 및 영양상태에 관한 통계생산과 국민건강증진 종합계획(Health plan)의 목표지표로서 WHO, OECD 등 국제기구에 보건지표를 제공하기 위하여 시행되고 있다(KCDC, 2010). KNHANES에 관한 자세한 내용은 부록 1로 첨부하였다.

전체 18,546건에서 65세 이상 만성질환을 가진 노인 자료는 2,991건이었으며, 이중 결측 및 이상 자료를 제외한 총 716건을 최종 연구대상으로 이용하였다.

본 연구는 EQ-5D 항목 및 추가로 시행된 검진조사항목(구강 검사, 골밀도 검사, 이비인후과 검사, 안검사)이 모두 포함되어 공통된 자료구조를 가진 2008~2010년에 실시된 4기 및 5기 조사자료를 1차 대상($n=18,546$)으로 하였다.

65세 이상의 노인을 대상으로 질병이환에 관한 설문에서 고혈압, 당뇨병, 간경변증, 뇌졸중 등의 만성질환에 해당되는 경우 유병여부 ‘있음’, 의사진단 여부 ‘있음’, 그리고 만성질환 진단시기가 ‘최소 1년 이상 경과’한 경우를 모두 만족하는 자료를 2차 대상($n=2,991$)으로 선별하였다. 이후 결측 및 이상 자료가 포함된 경우를 제외하고 3차 최종 대상($n=716$)을 선정하였다.

본 연구의 자료추출 및 분석과정은 다음과 같다(그림 11).

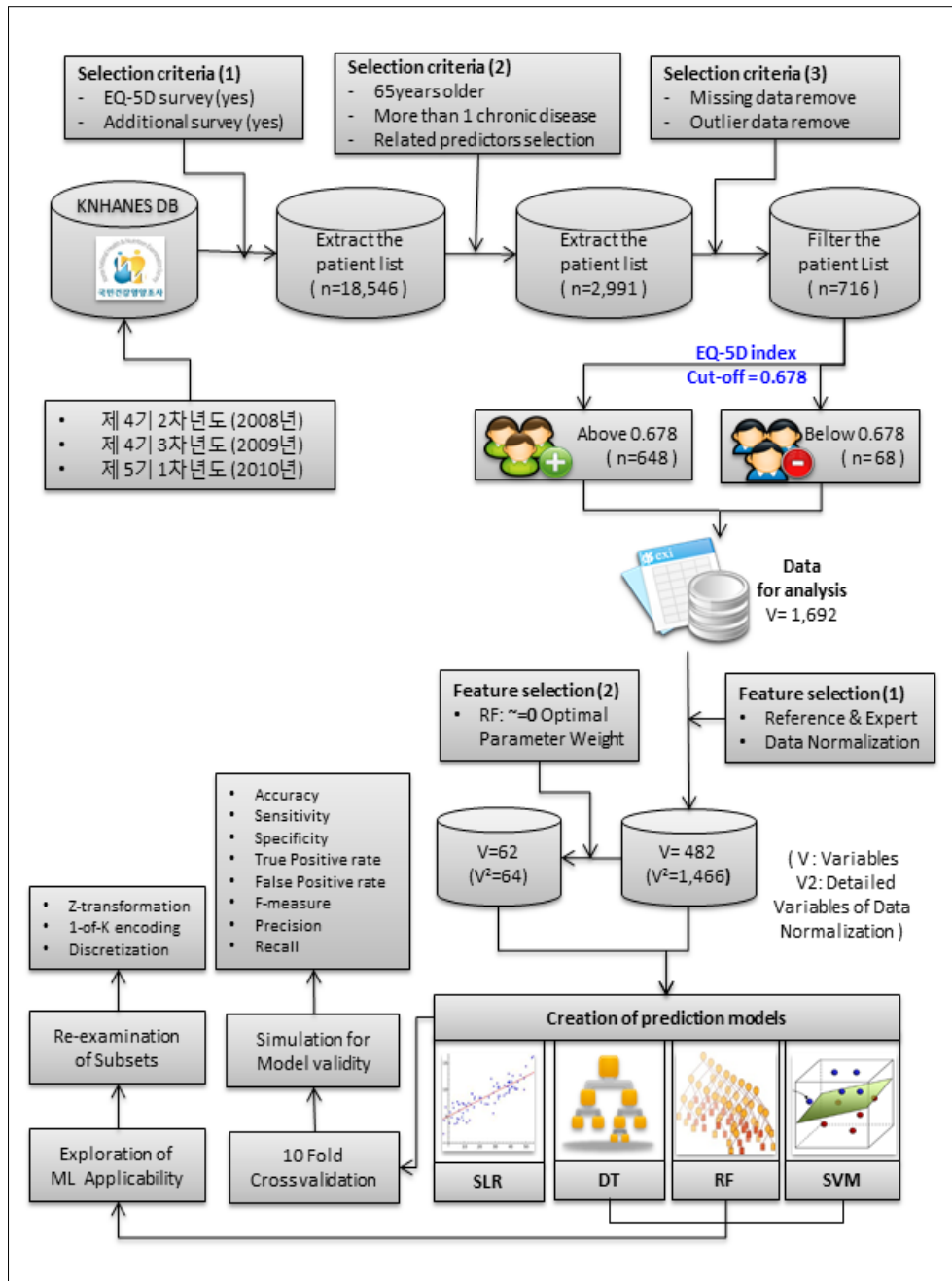


Figure 11. Procedure for extracting data and analysis

3. 연구 도구

3.1 EQ-5D index

KNHANES에서는 HRQoL 측정도구로 1987년에 설립된 EuroQol 그룹에서 다차원적 선호도에 근거한 HRQoL(multidimensional preference based Health Related Quality of Life)을 측정하기 위해 개발한 EQ-5D (Group, 1990)가 사용되었다.

EQ-5D는 선호도 및 일반적 건강상태를 측정하기 위해 널리 사용하고 있는 도구 중 하나이다. EQ-5D는 건강관련 삶의 질 상태를 기술하고 가치화하기 위해 개발한 도구로 운동능력(Mobility), 자기관리(Self-care), 일상 활동(Usual activity), 통증과 불편감(Pain and disability), 그리고 불안과 우울(Anxiety and depression)의 5개 항목에 대해서 현 상태를 ‘지장이 없음’, ‘다소 지장이 있음’, ‘할 수 없음’의 3개 수준 중 하나로 평가하도록 되어 있다. EuroQol Group의 승인을 받아 조사된 EQ-5D 조사표를 부록 2로 첨부하였다.

설문 응답자가 자신의 상태를 표현하면, 일반 인구집단에서 모든 EQ-5D 상태에 대해 질 가중치를 구해놓은 가치평가세트(Valuation set)를 이용하여 각 EQ-5D 상태에 해당하는 질 가중치 값을 구할 수 있는데 이를 EQ-5D index라고 한다. 5문항을 모두 1번(지장 없음)으로 응답한 경우 EQ-5D index는 ‘1.0’ 값이 주어진다. 반응변수 EQ-5D index는 -0.171과 1.0의 범위 내에 있으며, 시간교환법을 이용한 질가중치 측정값은 완전한 건강상태는 1, 가장 불완전한 상태는 -0.171에 해당된다. EQ-5D index 점수가 높을수록 HRQoL이 높다고 할 수 있다. EQ-5D 각 문항은 3개의 수준으로 측정되어, 가능한 건강수준의 조합은 총 $3^5=243$ 가지이다(Nam, et al., 2007). 243가지 조합에 따른 개별적인 가중치 추정값들은 부록 3으로 첨부하였다.

노인 만성질환자의 HRQoL 수준에 따라 건강관리 프로그램의 적용 여부

와 내용이 달라질 수 있다. 그러나 보건의료 정책측면에서 사회적 비용을 고려하였을 때 연속적으로 발생하는 모든 수준을 개별적으로 동시에 접근하기 보다는 취약집단을 선별하여 우선순위별로 중재를 하는 단계적인 접근방법이 요구되기 때문에 연속선상에 있는 HRQoL 척도를 그룹화하여 분류할 필요가 있다.

기존 연구에서는 HRQoL을 그룹으로 분류하는 cut-off point로 통계학적 관점에서 중위수(median)(Picavet & Hoeymans, 2004; van der Zanden et al., 2006) 또는 백분위수(percentile)(Joung, 2011) 등의 기준을 제시하였다. 반면 암환자를 대상으로 하는 대표적인 건강관련 삶의 질 측정도구인 EORTC-QLQ (European Organization for Research and Treatment of Cancer-quality of life questionnaires) 관련 연구들에 의하면 설문문항에 ‘꽤 그렇다’고 답변한 경우인 functional scores 66.7점을 임상적 측면에서 손상(impairment)의 역치(treshhold)로 판단하여 그룹을 양분하는 cut-off point 기준으로 제시하였다(Ahn et al., 2007; Stromgren et al., 2002).

따라서 본 연구에서는 중재가 필요한 HRQoL 위험군 예측모델의 개발을 위해 통계적인 분류기준보다는 중재 필요성 여부를 판단하기에 적합한 임상적인 분류기준에 따라 그룹을 분류하였다. 즉, EQ-5D 5문항의 점수를 합산하면 최소 5점(모든 문항에서 ‘지장 없음’을 선택한 경우)에서 최대 15점(모든 문항에서 ‘매우 심함’을 선택한 경우)이 나오게 된다. 이중 5문항 모두에 대해서 ‘다소 지장이 있다’고 대답한 경우는 합산점수의 중간 값인 10점에 해당되며, 이는 가중치를 반영한 EQ-5D index 값 0.677에 해당된다. 따라서 2개의 그룹으로 분류하기 위한 cut-off point를 EQ-5D index 값 0.678로 정하였으며, 이 값을 기준으로 0.678 이상인 그룹과 0.678 미만인 그룹으로 분류하였다. 위험군은 HRQoL이 상대적으로 저하되어 HRQoL 향상을 위한 중재가 필요한 EQ-5D index 0.678 미만인 그룹을 의미한다.

3.2 개인적 특성

(1) 연령

본 연구에서는 만 연령 65세 이상의 노인을 대상으로 하였다. 그러나 노년기의 확대에 의해 20~30년의 넓은 연령범위를 연령차이에 따른 노인의 욕구나 문제의 차이를 고려하여 살펴볼 필요가 있다. 따라서 KNHANES에서 개방형 질문으로 조사된 만나이를 연령대별로 전기노인(60~69세), 중기노인(70~79세), 그리고 후기노인(80세 이상)의 세 집단(최윤진, 2009)으로 분류하였다.

(2) 성별

KNHANES에서 제시한 ‘성별’ 항목으로서 남자는 ‘1’, 여자는 ‘2’로 나누어 가변수화된 자료를 이용하였다.

(3) 학력

KNHANES에서 제시한 ‘교육수준 재분류 코드’ 항목으로서 ‘초졸 이하’, ‘중졸’, ‘고졸’, ‘대졸 이상’으로 분류된 원자료를 사용하였다.

3.3 환경적 특성

(1) 소득

KNHANES에서 제시한 ‘소득 사분위수’ 항목으로서 ‘하’, ‘중하’, ‘중상’, ‘상’으로 구분된 원자료를 사용하였다. 소득은 균등화 소득에 근거하여 산출된 사분위수를 적용하였다(월평균 가구총소득/가구원수의 제곱근)(KCDC, 2010). 균등화 소득이란 가구특성(가구원수, 가구원의 인구학적 특성 등)이 다르면, 동일 후생수준 유지에 필요한 지출규모도 다르다는 점을 감안하여 조정된 가구소득을 의미한다.

KNHANES에서 개방형 질문으로 조사된 ‘월평균 가구총소득’ 항목의 분

석을 위하여 ‘0~50만원’, ‘51~85만원’, ‘86~200만원’, ‘201만원 이상’으로 분류하였다.

(2) 직업

KNHANES 자료에서 제시한 ‘직업재분류 및 실업/비경제활동 상태 코드’ 항목으로서 ‘관리자, 전문가 및 관련 종사자’, ‘사무종사자’, ‘서비스 및 판매 종사자’, ‘농림어업 숙련 종사자’, ‘기능원, 장치기계조작 및 조립종사자’, ‘단순노무종사자’, ‘무직(주부, 학생 등)’으로 구분된 원자료를 사용하였다.

(3) 가족

KNHANES 자료에서 제시한 ‘가구세대 구성코드’ 항목으로서 ‘1인 가구’, ‘부부(응답자+ 배우자)’, ‘부부+ 미혼 형제자매’, ‘응답자+ 미혼 형제자매’, ‘부부+ 기타 친인척’, ‘응답자+ 기타 친인척’, ‘상기를 제외한 모든 1세대 가구’, ‘부부+ 미혼자녀’, ‘편부+ 미혼자녀’, ‘편모+ 미혼자녀’, ‘부부+ 양친’, ‘부부+ 편부모’, ‘부부+ 자녀+ 부부의 형제 또는 자매’, ‘조부모+ 미혼 손자녀’, ‘편조부/편조모+ 미혼손자녀’, ‘상기를 제외한 모든 2세대 가구’, ‘부부+ 미혼자녀+ 양친’, ‘부부+ 미혼자녀+ 편부모’, ‘상기를 제외한 모든 3세대 이상 가구’로 구분된 원자료를 사용하였다.

3.4 생리적 요인

(1) 만성질환 종류

KNHANES에 조사된 만성질환으로 국한하였으며, 고혈압, 고지혈증, 뇌졸중, 협심증, 심근경색증, 당뇨병, 갑상선 질환, 신부전, 골관절염, 류마티스성 관절염, 골다공증, 폐결핵, 천식, 위암, 간암, 대장암, 유방암, 자궁경부암, 폐암, 기타암, B형 간염, C형 간염, 간경변증을 포함하였다.

(2) 만성질환 개수

KNHANES의 질병이환에 관한 설문에서 상기 만성질환에 해당되고, 유병여부 ‘있음’과 의사진단여부 ‘있음’으로 대답하였으며, 만성질환 진단시기가 최소 1년 이상 경과한 경우를 선별하였다. 각각의 질환을 1개로 간주하여 해당되는 모든 질환의 수를 합산하였다. 분석을 위하여 ‘1개’, ‘2개’, ‘3개’, ‘4개’, ‘5개 이상’으로 구분하였다.

(3) 만성질환 합산기간

KNHANES의 질병이환에 관한 설문에서 상기 만성질환 진단 후 소요기간을 합산하였다. 소요기간은 만연령에서 질환별 진단시기의 나이를 뺀 기간으로 산정하였다. 분석을 위하여 ‘1~5년’, ‘6~14년’, ‘15~30년’, ‘31~60년’, ‘61년 이상’으로 구분하였다.

(4) 체질량지수

BMI는 체중을 신장의 제곱으로 나눈 뒤 백분율로 측정한 수치로 KNHANES에서 제시한 원자료를 이용하였다. BMI는 대한비만학회의 제안에 따라 18.5~24.9kg/m²이면 ‘정상체중’, 18.5 미만은 ‘저체중’, 25 이상은 ‘비만’으로 분류하였다.

3.5 증상경험

(1) 우울

KNHANES의 정신건강에 관한 설문에서 ‘2주 이상 연속 우울감 여부’ 항목으로서 ‘예’, ‘아니오’, ‘모름’으로 분류된 원자료를 사용하였다.

(2) 불편감

KNHANES에서 개방형 질문으로 조사된 ‘최근 2주간 불편감일수’ 항목을 분석을 위하여 ‘0일’, ‘1~7일’, ‘8~14일’로 분류하였다.

3.6 일반적 건강 지각

(1) 주관적 건강상태

KNHANES에서 ‘평소에 건강은 어떻다고 생각하십니까?’라는 질문에 답변한 ‘주관적 건강상태’ 항목으로서 ‘매우 좋음’, ‘좋음’, ‘보통’, ‘나쁨’, ‘매우 나쁨’으로 분류된 원자료를 사용하였다.

3.7 건강증진행위

(1) 활동량

KNHANES에서 조사된 ‘1주일간 걷기일수’ 항목으로서 ‘전혀 하지 않음’, ‘1일’, ‘2일’, ‘3일’, ‘4일’, ‘5일’, ‘6일’, ‘7일(매일)’로 분류된 원자료를 사용하였다.

(2) 흡연

KNHANES에서 개방형으로 조사된 ‘과거흡연자 하루 평균 흡연량’ 항목으로서 원자료를 사용하였다.

(3) 음주

KNHANES에서 조사된 ‘1년간 음주빈도’ 항목으로서 ‘최근 1년간 전혀 마시지 않았다’, ‘월1회 미만’, ‘월1회 정도’, ‘월2~4회’, ‘주2~3회 정도’, ‘주4회 이상’으로 분류된 원자료를 사용하였다.

KNHANES에서 조사된 ‘음주로 인한 상해여부’ 항목으로서 ‘없었다’, ‘과거에는 있었지만 최근 1년 동안에는 없었다’, ‘최근 1년 동안에 그런 적이 있었다’로 분류된 원자료를 사용하였다.

4. 자료 정규화

자료 정규화(Data Normalization)란 자료의 손실이나 불필요한 정보의 도입 없이 데이터의 일관성, 최소한의 데이터 중복, 최대의 데이터 안정성 확보를 위한 안정적 자료구조로 변환하는 기법을 의미한다. 본 연구에서 적용한 자료 정규화 방법들은 다음과 같다.

4.1 Z-변형(Z-transformation)

Z-변형(Z-transformation)이란 기준이 다른 두 가지 이상의 표준편차의 자료를 비교하기 쉽도록 평균과 분산이 각각 0과 1이 되도록 변환하는 과정을 말한다(Cheadle, Vawter, Freed, & Becker, 2003). 즉, 서로 다른 척도의 자료들을 동일한 척도인 Z-score로 변형하여 비교할 수 있게 하는 방법으로 척도의 차이에서 발생하는 오차를 줄이기 위해 사용하게 된다. 본 연구에서 Z-변형을 위해 기본값으로 autoscale이 포함된 MATLAB function 'Svmtrain'을 사용하였다.

4.2 이진화(1-of-K encoding scheme coding)

연속변수(Continuous variable)의 숫자는 크기에 따른 의미가 있지만, 명목변수(Nominal variable)의 숫자는 의미가 없다. 예를 들면 성별의 경우 남자=1, 여자=2라고 한다면 사람은 이 숫자가 의미가 없는 약속이라는 것을 이해하지만 기계적으로는 여자는 남자보다 2배의 value를 가진다고 잘못 해석할 수 있기 때문에 1-of-K encoding scheme coding이 필요하다(Bishop, 2006).

1-of-K encoding scheme coding은 남자 0 1 True, 여자 1 0 False 라는 값으로 의미를 부여하기 때문에 계산상의 오차를 없애주고 실험의 결과 값이 잘 나오도록 성능을 높여주는 역할을 한다. 따라서 본 연구에서는 기존

독립변수 중 다항변수(Multi-nominal Data)는 1-of-K encoding scheme coding하여 변환하였다.

4.3 이산화(Discretization)

ML 기법에 적용되는 실세계 데이터의 속성(attribute)은 연속형(continuous)과 범주형(categorical)의 혼합된 형태(Mixed-type)를 지니게 된다. 그러나 대부분의 ML 알고리즘은 한 가지 형태의 데이터만을 다룰 수 있기 때문에, 기계학습 알고리즘에 적용시키기 위해서는 데이터 속성의 형 변환이 요구된다. 일반적으로 범주형 값을 연속형으로 변환시키는 문제의 복잡성과 범주형 값이 분류규칙(classification rule)을 도출하기에 더 용이하다는 장점으로 인해 연속형을 범주형으로 변환하는 기법인 이산화(discretization) 방법을 많이 사용한다(Witten & Frank, 2005). 원래 데이터의 정보를 유지하면서 동시에 일반화된 대표값을 산출할 수 있는 적절한 수준의 이산화 구간을 찾는 것이 이산화 알고리즘의 주된 목적이다(Luengo, Saez, Lopez, & Herrera, 2012).

본 연구에서는 일부 연속변수를 순서변수, 명목변수로 변형하여 부분적인 재실험을 통해 성능을 비교하였다. 이산화를 위해서 해당 변수의 통계적 특징(분포, 평균 등)과 전문가 지식(정상 범위 등)을 활용하였다.

4.4 새로운 변수 생성

실수값을 갖는 연속변수(Continuous variable)는 기존 독립변수를 조합하여 새로운 독립변수를 생성하여 의미를 확장하였다. 예를 들어 ‘와병일수’, ‘입원기간’ 등의 변수를 합산처리하거나 ‘젊은시절 신장’과 ‘현재 신장’과의 차이를 계산해서 ‘신장 변화량’이라는 새로운 변수를 생성하였다.

선행연구와 전문가 의견에 근거하여 노인 만성질환자들의 HRQoL과 상

관관계에 있을 것이라 예상되는 항목을 독립변수로 선정하였고, 결측 및 이상 자료를 제거하고, 자료 정규화 과정을 거쳐 1,692개 변수를 최종 498개 (상세 1,482개)의 변수로 조정하였다. 그 결과를 변수 정규화 방법표로 정리하였으며, 그 내용의 일부분을 살펴보면 다음과 같다(표 4). 변수 정규화 방법에 대한 전체 내용은 부록 4로 첨부하였다.

Table 4. Method table of data normalization (sample)

Variables	Explanation	Classification	Method	Misc
Edu	Education level	1. Below elementary 2. Middle school graduate 3. High school graduate 4. Above college graduate	1-of-K encoding	1→0001 2→0010 3→0100 4→1000
marri_1	Marriage	1. Married 2. Single 9. unknown	1-of-K encoding	1→0001 2→0010 3→0100
CD_N	Chronic diseases number (Prevalence & Diagnosis & More than 1year)	□□ea	Real Data	Newly created
CD_D	Chronic diseases total duration (Now - Diagnosis age)	□□□year	Real Data	Newly created
CD_DM	Chronic diseases individual duration (Total duration/ea)	□□year	Real Data	Newly created
HE_ht_cg	Height variation (Young - Now height)	□□.□cm	Real Data	Newly created

5. 자료 분석

자료분석 및 모형개발을 위해 SPSS version 20.0 프로그램(IBM, 2012)과 MATLAB (Matrix Laboratory) Revised for Version 7.14 (Release 2012a) 프로그램을(MathWorks, 2012)을 사용하였다. MATLAB은 MathWorks사에서 개발한 공학용 소프트웨어로 알고리즘 개발, 데이터 분석, 시각화 및 수치 연산을 위한 프로그래밍 환경을 제공한다.

SPSS로 연구 대상자의 변수에 대한 기술통계, X^2 -test, 단계적 로지스틱 회귀분석(Stepwise Logistic Regression: 이하 SLR)을 실시하였으며, 유의수준은 $p < 0.05$ 로 통계적 유의성을 검증하였다. MATLAB을 이용하여 DT, RF, SVM 모델을 구축하였다.

구축된 모델의 타당도 검증은 일반적으로 사용되는 10배수 교차타당성(10 fold cross-validation) 방법을 이용하였다. 즉, 검증용 데이터를 10개의 subset으로 무작위 배정하여 돌아가면서 9개의 subset 데이터는 모델 구축에 사용하고, 나머지 1개의 subset 데이터는 구축된 모델을 평가하는데 사용하였다. 본 연구에서는 총 10회의 검증 결과를 평균한 값을 결과로 산출하였다.

구축된 모델평가는 각 모델의 10 fold cross-validation에 대한 분류행렬(Confusion matrix: 이하 CM)을 통해 연구설계 시 설정한 실제값과 예측값을 비교한 분류정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity), 진양성율(True Positive rate), 위양성율(False Positive rate) 값을 각각의 모형별로 비교 확인하였다.

CM은 통계모델을 평가하기 위한 표준도구로 혼동행렬이라고도 한다. 행렬에 대한 행은 모델의 예측값을 나타내고 열은 실제값을 나타낸다. 분석에는 진양성(True Positives: 이하 TP), 위양성(False Positives: 이하 FP), 진음성(True Negatives: 이하 TN), 위음성(False Negatives: 이하 FN)의 4가지 범주가 사용되며, CM은 잘못된 예측의 영향을 간편하게 파악하고 분석할 수 있으므로 예측의 결과를 평가할 수 있는 중요한 도구이다. 이 행

렬의 각 셀에 나와있는 합계와 비율을 통해 모델이 정확하게 예측한 빈도를 빠르게 확인할 수 있다.

정확율(Precision)은 검색한 자료 중에서 적합한 자료의 비율, 즉 검색의 정확성을 측정하는 지표이며, 재현율(Recall)은 적합한 자료를 얼마나 많이 검색했느냐, 즉 적합한 자료를 찾아내는 능력을 나타내는 지표로 반비례 관계에 있기 때문에 두 개념을 동시에 반영하는 지표인 F-score 값을 함께 측정하였다.

각각의 공식은 아래와 같다.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$True\ Positive\ rate = \frac{TP}{(TP + FN)}$$

$$False\ Positive\ rate = \frac{FP}{(FP + TN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$F - score = \frac{2TP}{(2TP + FP + FN)}$$

V. 연구 결과

1. 대상자의 일반적 특성

KNHANES의 원시자료를 대상으로 순차적으로 추출하였으며, 우선적으로 EQ-5D 설문문항이 포함된 공통된 자료구조를 가진 대상을 추출하였다. 이 결과로 1차 자료(n=18,546)를 얻었으며, 여기에서 만 65세 이상 노인(n=3,786)에 대한 자료를 추출하였다(표 5).

Table 5. Case selection result (unit: person)

Year	Number	More than 65 age (%)
2010	7,043	1,378 (19.6)
2009	7,920	1,627 (20.5)
2008	3,583	781 (21.8)
Total	18,546	3,786 (20.4)

만성질환 여부는 KNHANES 자료에서 조사된 23개 상병(고혈압, 고지혈증, 뇌졸중, 심근경색증, 협심증, 당뇨병, 갑상샘 질환, 신부전, 골관절염, 류마치스성 관절염, 폐결핵, 천식, 위암, 간암, 대장암, 유방암, 자궁경부암, 폐암, 기타암, B형간염, C형간염, 간경변증)의 유병여부 ‘있음’과 의사진단 여부 ‘있음’, 그리고 만성질환 진단시기가 ‘최소 1년 이상 경과’한 경우 만성질환이 있는 것으로 정의하였다.

진단시기가 조사되지 않아 기간이 명시되지 않은 골다공증과 이환항목으로 조사되지 않은 만성폐쇄성폐질환 상병은 제외하였다. 만 65세 이상 노인을 대상으로 하는 자료(n=3,786)에서 만성질환 개수가 1개 이상인 경우에 해당되는 2차 자료(n=2,991)를 추출하였다.

만 65세 이상 노인을 대상으로 만성질환 여부에 따른 HRQoL 수준을 확

인한 결과, 만성질환을 가지고 있는 그룹이 만성질환이 없는 그룹에 비해 EQ-5D index 평균값이 낮았으며, 통계적으로 유의한 차이를 나타내었다 (표 6).

Table 6. EQ-5D index by chronic disease groups

Chronic Disease	N (%)	Mean	SD	t	p-value
Yes	2,991 (79)	0.825	0.197	-8.94	0.000*
No	795 (21)	0.895	0.151		
Total	3,786 (100)				

* $p < 0.05$, SD: Standard deviation

만성질환을 가진 만 65세 이상 노인들에 해당되는 2차 자료(n=2,991)에서 설문항목의 결측 및 이상 자료를 제외한 3차 자료(n=716)를 최종 대상으로 선정하였다. 본 연구 대상자의 일반적 특성은 표 7과 같다.

Table 7. Descriptive statistics for observed variables (n=716)

Variables	Category	n (%)	Mean±SD
Age(yrs)			70.6±4.7
Gender	Male	379 (52.9)	
	Female	337 (47.1)	
Income	Low	136 (19.0)	
	Medium-low	174 (24.3)	
	Medium-high	185 (25.8)	
	High	221 (30.9)	
Education level	Below elementary	420 (58.7)	
	Middle school graduate	118 (16.5)	
	High school graduate	109 (15.2)	
	Above college graduate	69 (9.6)	
Marital status	Spouse (attached)	536 (74.9)	
	Spouse (separated)	5 (0.7)	
	Widowed	157 (21.9)	
	Divorce	13 (1.8)	
	Other (non-applicable)	5 (0.7)	
Job	Managers, professionals	6 (0.8)	
	White-collar	5 (0.7)	
	Service, sales	31 (4.3)	
	Agriculture, fishing	146 (20.4)	
	Functions, device	20 (2.8)	
	Simple labor	57 (8.0)	
	Unemployed	451 (63.0)	
Health insurance	Local health insurance	234 (32.7)	
	Workplace insurance	442 (61.7)	
	Medical Aid	33 (4.6)	
	Other (uninsured)	7 (1.0)	

대상자의 평균 연령은 만 70.6세였으며, 소득은 하에 속하는 경우가 19%, 상에 속하는 그룹이 30.9%였다. 초등학교 졸업이하 학력이 58.7%로 가장 많았으며, 결혼상태는 배우자와 동거하는 형태가 74.9%, 사별이 21.9%에 해당되었다. 직업은 무직이 63%로 가장 많았고, 농림 어업에 종사하는 경우가 20.4%를 차지하였다. 건강보험의 종류에서 직장가입자가 61.7%에 해당되었는데, 이는 무직이 63%라는 사실을 감안해보면 직장가입자인 자녀의 피부양자로서 직장의료보험에 가입된 것으로 생각된다.

2. 대상자의 건강관련 삶의 질 수준

본 연구에서는 HRQoL 그룹을 종속변수로 정하였으며, 이는 EQ-5D index 점수를 기준으로 2개의 그룹으로 분류한 것이다. EQ-5D index 점수를 cut-off point인 0.678을 기준으로 0.678 이상인 그룹은 '1'로 0.678 미만인 위험군의 경우는 '0'으로 분류하였다.

대상자들의 EQ-5D index 분포를 살펴보면 오른쪽으로 치우친 비대칭 분포(skewed distribution)를 나타내었으며, 적색선은 HRQoL 그룹을 분류하는 cut-off point인 0.678을 표시하였다(그림 12).

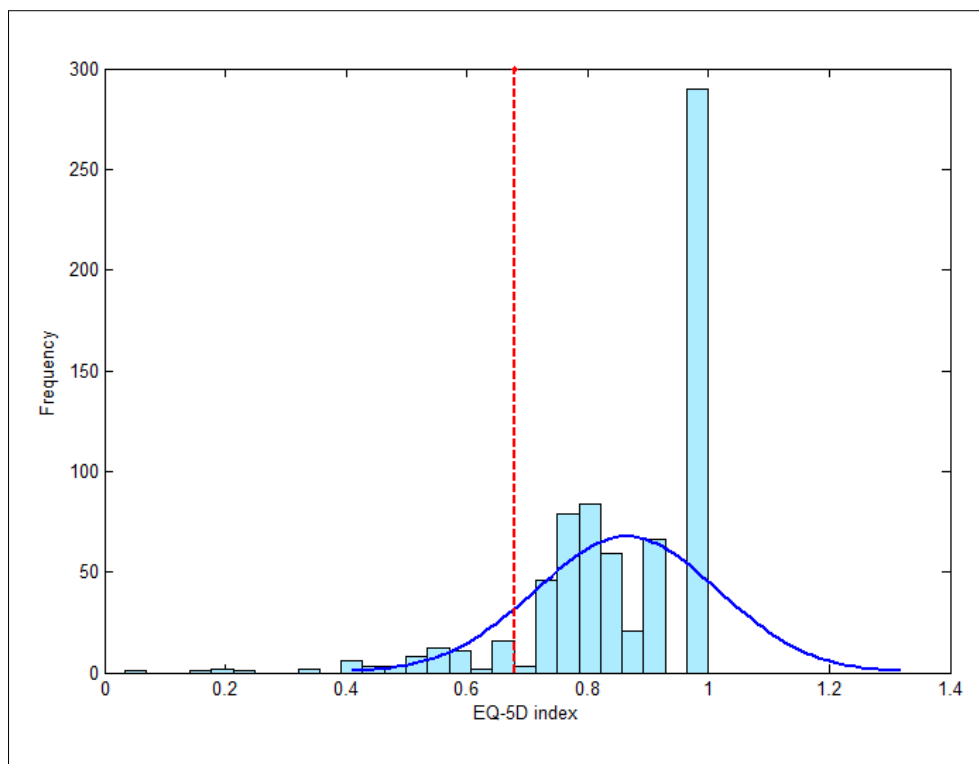


Figure 12. Distribution of EQ-5D index

HRQoL EQ-5D index 0.678 이상인 그룹이 0.678 미만인 그룹에 비해 약 10배나 많은 빈도를 나타내었으며, HRQoL 그룹별 EQ-5D index 평균은 유의한 차이를 나타내었다(표 8).

Table 8. EQ-5D index by groups

Group	N (%)	Mean	SD	t	p-value
Below 0.678	68 (9.5)	0.530	0.139	-27.56	0.000*
Above 0.678	648 (90.5)	0.900	0.101		
전체	716 (100)	0.865	0.151		

* $p < 0.05$, SD: Standard deviation

대상자의 일반적 특성에 따른 HRQoL 수준의 차이를 살펴보면 다음과 같다(표 9).

Table 9. Differences of health-related quality of life by general characteristics

Variables		M (SD)	EQ-5D index		X ²	p-value
			Below 0.678	Above 0.678		
Age	Former elderly (60s)	0.903 (0.098)	34	322	1.70	0.427
	Middle elderly (70s)	0.900 (0.103)	27	285		
	Latter elderly (80s)	0.873 (0.110)	7	41		
Gender	Male	0.926 (0.095)	28	351	4.17	0.041*
	Female	0.868 (0.995)	40	297		
Education level	Below elementary	0.879 (0.102)	57	363	20.17	0.000*
	Middle school graduate	0.914 (0.095)	5	113		
	High school graduate	0.925 (0.096)	5	104		
	Above college graduate	0.947 (0.090)	1	68		
Monthly income (10,000 won)	0~50	0.875 (0.106)	35	180	23.31	0.000*
	51~85	0.890 (0.097)	16	129		
	86~200	0.922 (0.091)	13	174		
	≥201	0.910 (0.103)	4	165		
Job	Professionals	0.970 (0.075)	0	6	5.39	0.495
	White-collar	0.917 (0.116)	0	5		
	Simple labor	0.868 (0.137)	4	53		
	Unemployed	0.862 (0.159)	44	407		
Spouse	Attached	0.879 (0.146)	44	492	4.12	0.042*
	Separated	0.822 (0.157)	24	156		
Family	1 person households	0.810 (0.186)	17	77	18.90	0.274
	Couple, single child, lone-parent	0.865 (0.148)	2	35		
	Couple, single child	0.877 (0.124)	5	60		
	Couple	0.879 (0.150)	30	335		
	Grandparents, single grandchildren	0.914 (0.112)	2	35		

Variables		M (SD)	EQ-5D index		X ²	p-value
			Below 0.678	Above 0.678		
Chronic diseases number	1	0.917 (0.096)	17	273	11.98	0.017*
	2	0.902 (0.993)	29	213		
	3	0.871 (0.103)	16	102		
	4	0.854 (0.102)	6	36		
	5~8	0.856 (0.114)	0	23		
Chronic diseases total duration (yrs)	1~5	0.920 (0.096)	14	171	1.59	0.810
	6~14	0.896 (0.101)	18	162		
	15~30	0.895 (0.101)	20	156		
	31~60	0.885 (0.106)	13	129		
	≥61	0.887 (0.103)	3	30		
Body mass index	Underweight (<18.5)	0.912 (0.107)	4	23	1.56	0.458
	Normal (18.5~24.9)	0.902 (0.099)	45	408		
	Obesity (≥25)	0.895 (0.104)	19	217		
Depression	Yes	0.869 (0.104)	30	102	32.96	0.000*
	No	0.906 (0.100)	38	546		
Discomfort days	0	0.930 (0.091)	11	387	61.82	0.000*
	1~7	0.877 (0.095)	7	82		
	8~14	0.844 (0.100)	50	179		
Sick days	0	0.882 (0.131)	43	586	74.17	0.000*
	1~7	0.792 (0.196)	13	53		
	8~14	0.707 (0.191)	2	4		
	15~30	0.538 (0.213)	10	5		
Perceived health status	Very good	0.961 (0.076)	1	30	105.7	0.000*
	Good	0.930 (0.092)	5	208		
	Fair	0.923 (0.090)	3	176		
	Poor	0.850 (0.986)	30	188		
	Very poor	0.840 (0.098)	29	46		
Drinking	Recently no	0.908 (0.098)	19	134	5.59	0.470
	Less than once a month	0.885 (0.120)	9	78		
	About once a month	0.849 (0.098)	3	43		

* $p < 0.05$ (Fisher's exact test result)

연령대에 따른 EQ-5D index 평균값은 60대가 가장 높았으며, 연령이 증가할수록 증가되었으나, HRQoL 그룹별 차이는 유의하지 않았다.

성별에 따른 EQ-5D index 평균값은 남자가 여자에 비해 높았으며, HRQoL 그룹별 차이는 유의하였다.

학력별 EQ-5D index 평균값은 대졸이상이 가장 높았으며, 학력이 높을수록 증가하였고, HRQoL 그룹별 차이는 유의하였다.

월평균 가구총소득의 경우 EQ-5D index 평균값은 소득이 증가할수록 높아지는 경향이나, 201만원 이상 그룹보다 80~200만원 그룹이 더 높았으며, HRQoL 그룹별 차이는 유의하였다.

직업재분류 항목의 경우 EQ-5D index 평균값은 관리자, 전문가 그룹이 가장 높고, 무직이 가장 낮았으나, HRQoL 그룹별 차이는 유의하지 않았다.

배우자 동거여부의 경우 동거 그룹이 별거, 이혼, 사별 등의 비동거 그룹에 비해 EQ-5D index 평균값이 높았으며, HRQoL 그룹별 차이는 유의하였다.

가구세대 구성코드 항목의 경우 EQ-5D index 평균값은 1인 가구가 가장 낮았고, 부부(응답자+ 배우자)인 경우보다 조부모+ 미혼 손자녀 그룹이 더 높았으나, HRQoL 그룹별 차이는 유의하지 않았다.

만성질환 개수가 증가할수록 EQ-5D index 평균값도 대체로 저하되는 경향이 있으나, 4개 그룹보다는 5~8개 그룹이 더 높았고, HRQoL 그룹별 차이는 유의하였다.

만성질환 합산기간이 증가할수록 EQ-5D index 평균값이 저하되는 경향이 있으며, 31~60년 그룹보다는 61년 이상 그룹이 약간 높았으나, HRQoL 그룹별 차이는 유의하지 않았다.

만성질환 개수와 만성질환 합산기간이 증가할수록 EQ-5D index가 계속 저하되는 것이 아니라 어떤 시점을 기준으로 다시 상승하는 경향이 나타나는데 이는 만성질환에 적응하고 대응능력이 강화되었기 때문이라고 볼 수 있다.

체질량지수에 있어서 비만인 경우 EQ-5D index 평균값이 가장 낮았으

나, HRQoL 그룹별 차이는 유의하지 않았다.

우울감이 없는 경우, 불편감 일수가 적을수록, 외병일수가 적을수록, 주관적 건강상태가 좋을수록 EQ-5D index 평균값이 높았으며, HRQoL 그룹별 차이는 유의하였다.

음주의 경우 전혀 마시지 않는 경우가 가장 높은 EQ-5D index 평균값을 나타내었으나, HRQoL 그룹별 차이는 유의하지 않았다.

3. 영향요인 분석

노인 만성질환자의 HRQoL에 영향을 미치는 요인을 파악하기 위해서 기계학습 기법인 DT, RF, SVM과 단계적 로지스틱 회귀분석을 실시하였다.

3.1 Decision Tree model

DT 적용결과는 그림 13과 같으며, 나무의 가장 상단에 위치하는 뿌리마디(root node)는 ‘활동제한사유(비만)’ 변수가 해당되었다. 그 다음으로는 ‘골관절염 소요기간’, ‘최근 2주간 불편감일수’, ‘뇌졸중 진단시기’, ‘골다공증 유병여부’, ‘골관절염 진단시기’, ‘주관적 건강상태’ 등이 중요변수로 확인되었다.

DT 적용결과 확인된 중요변수는 최종 26개로 표 10과 같다.

Table 10. Important variables in DT

Variables	Explanation
LQ4_21	Restriction of activity (Obesity)
DM2_ag_d	Osteoarthritis duration
D_2_wk	Last 2 weeks of discomfort days
DI3_ag	Stroke diagnosis time
DX_OST_FN	Prevalence whether osteoporosis (Femoral neck)
DM2_ag	Osteoarthritis diagnosis time
D_1_1	Perceived health status
ainc	Monthly income
DC6_ag	Lung cancer diagnosis time
BP5	Continuous depression for more than two weeks
BP6_10	1 year suicidal ideation
HE_glu	Fasting blood glucose
T_Q_CR2	Rhinitis week duration
HE_fst	Fasting hours
HE_wc	Waist circumference
HE_insulin	Insulin (μ IU/mL)
O_DMFTP	Number of permanent teeth caries experience
age	Age
BS3_2	Current smokers, cigarettes per a day
DM3_ag_d	Rheumatoid arthritis duration
O_DTP	Number of caries permanent teeth
MO4_1_01	Outpatient utilization reasons
apt_t	Housing types
DI2_ag	Hyperlipidemia diagnosis time
DJ2_ag_d	Tuberculosis duration
DX_S_Ts_A	Lumbar T-score (Asian standards)

3.2 Random Forest model

RF는 variable importance를 통해 모형의 중요변수를 추정할 수 있다. 이에 RF 적용결과 최적변수 가중치(Optimal Parameter Weight: 이하 OPW)를 얻을 수 있었으며, OPW의 결과값이 0이 아닌 변수(V2=64)들을 표 11과 같이 Heat map으로 표현하였다. 가장 중요도가 높은 변수는 ‘만성 질환 개수’에 해당되며, ‘인슐린 혈액검사’가 그 다음으로 높은 가중치를 나타내었다. 그 외에도 ‘2주이상 연속 우울감 여부’, ‘최근 2주간 불편감 일수’, ‘고혈압 진단시기’, ‘당뇨병 소요기간’, ‘당뇨병 진단시기’, ‘1주일간 걷기 일수’, ‘주관적 건강상태’, ‘체질량 지수’, ‘활동제한 사유: 비만’ 등도 중요 변수에 해당되었다.

Table 11. Heat map of important feature variables in RF

Variables	Explanation	Weight
CD_N	Chronic diseases number	1.731
BP5	Continuous depression for more than two weeks	0.999
DI1_ag	Hypertension diagnosis time	0.999
DE1_ag_d	Diabetes duration	0.892
D_2_wk	Last 2 weeks of discomfort days	0.726
DX_F_Ts_A	Femoral bone T-score(Asian standards)	0.577
ainc	Monthly income	0.577
DI3_ag	Stroke diagnosis time	0.577
DI6_ag	Angina diagnosis time	0.577
DM3_ag_d	Rheumatoid arthritis duration	0.577
BE3_22_23	Moderate physical activity duration	0.577
HEfst	Fasting hours	0.577
HE_wc	Waist circumference	0.577
HE_BMI	Body mass index (BMI)	0.577
MO7_2	Copayments for outpatient	0.577
LQ1_mn	Recently 1 month sick days	0.577
edu	Education level	0.577
occp	Job	0.577

Variables	Explanation	Weight
DX_OST_LS	Prevalence whether osteoporosis (Lumbar)	0.577
D_1_1	Perceived health status (Good)	0.577
D_2_1	The last 2 weeks of discomfort (No)	0.577
DI4_pt	Myocardial infarction, angina treatment (No)	0.577
DI6_pt	Aangina treatment (Not applicable)	0.577
LQ4_22	Restriction of activity (Old-age)	0.577
BD1_11	1 year frequency of drinking (No)	0.577
BE3_31	1 week walking days (6.5 days)	0.577
HE_DM	Prevalence of diabetes	0.577
MO3_2_04	Outpatient utilization reasons (Not applicable)	0.577
O_PROS_L	Mandibular prosthesis status (1 dentures fixed)	0.577
DM2_ag_d	Osteoarthritis diagnosis time	0.326
DX_FN_Ts_A	Femoral neck T-score (Asian standards)	0.013
HE_ht_cg	Height variation (Young – Now height)	0.004
N_DIET_WHY	Diet reasons (Weight control)	-0.009
MO4_1_01	Outpatient utilization reasons (Cold)	-0.013
DM2_ag	Osteoarthritis diagnosis time	-0.024
HE_sbp_tr	End-systolic pressure (Calibrated value)	-0.134
O_DMFTP	Number of permanent teeth caries experience	-0.320
HE_dbp_tr	End-diastolic pressure (Calibration value)	-0.401
CD_D	Chronic diseases total duration	-0.533
DI2_ag_d	Hyperlipidemia duration	-0.577
DI3_ag_d	Stroke duration	-0.577
DE1_ag	Diabetes diagnosis time	-0.577
BS6_3	Former smokers cigarettes per a day	-0.577
BD5_R	Affect the frequency of daily life due to drinking	-0.577
BE3_11_R	Vigorous physical activities for 1 week	-0.577
genertn	Family (Couple, unmarried children, lone-parent)	-0.577
D_1_1	Perceived health status (Very poor)	-0.577
DM1_dg	Arthritis doctor diagnose (Yes)	-0.577
DF2_dg	Depression doctor diagnose (No)	-0.577
LQ4_07	Restriction of activity (Diabetes)	-0.577
LQ4_21	Restriction of activity (Obesity)	-0.577
graduat	Graduation level of education (Dropout)	-0.577

Variables	Explanation	Weight
EC1_2	Unemployed reason (Did not feel the need)	-0.577
EC_pedu_1	Father education level (Below elementary)	-0.577
BO2_1	1 years of weight control (No)	-0.577
BD1_11	1 year frequency of drinking (2-3 per week)	-0.577
BD7_3	Injury caused by drinking (No)	-0.577
BE3_31	1 week walking days (3.2 days)	-0.577
HE_HP_tr	Hypertension prevalence	-0.577
T_Q_CR3	Rhinitis symptom duration (More than 1 month)	-0.577
MP4_2_01	Pharmacy purchases a prescription drug (No)	-0.577
MP4_2_03	Pharmacy purchases a prescription drug (Yes)	-0.577
MP4_6_01	Pharmacy purchases other (Yes)	-0.577
HE_insulin	Insulin (μ IU/mL)	-1.000

3.3 Support Vector Machine model

SVM 적용결과 최적의 성능을 나타내는 중요변수들은 RF의 OPW값이 0 이상인 총 32개의 변수들이 해당되었다.

3.4 Stepwise Logistic Regression model

ML의 3가지 방법에서 선정된 중요변수들과 문헌고찰을 통해 중요하다고 판단되는 변수들을 합하여 총 74개를 Input data로 활용하였다. 단계적 로지스틱 회귀분석 적용결과 총 8개의 변수가 유의한 변수로 나타났으며, 그 결과는 표 12와 같다.

Table 12. Result of stepwise logistic regression

Variables	B	S.E	<i>p</i> -value	Exp(B)	95% CI for Exp(B)	
					Lower	Upper
Continuous depression for more than 2 weeks (1=Yes, 2=No)	1.018	0.330	0.002	2.767	1.449	5.285
Myocardial infarction, angina treatment (1=Yes, 0=No)	-0.312	0.150	0.038	0.732	0.545	0.983
Outpatient utilization reasons (1=Disease, 2= Accident, poisoning, 3=Others)	0.165	0.058	0.004	1.179	1.053	1.321
Monthly income (10,000won)	0.004	0.002	0.007	1.004	1.001	1.007
Perceived health status (1=Very good~5=Very poor)	-0.490	0.193	0.011	0.612	0.419	0.894
Recently 1 month sick days	-0.068	0.025	0.007	0.934	0.889	0.982
Last 2 weeks of discomfort days	-0.086	0.028	0.002	0.918	0.870	0.969
Restriction of activity (Old age) (1=Yes, 0=No)	0.167	0.055	0.002	1.181	1.061	1.316
Constant	3.413	1.617	0.035	30.367		

3.5 Comparison of influencing factors

노인 만성질환자의 HRQoL에 영향을 미치는 중요변수들을 연구모형의 영향요인 분류에 맞추어 정리하면 다음과 같다(표 13).

Table 13. Important variables of four models

Factors	Variables	DT	RF	SVM	SLR
Characteristics of the Individual	Age	0			
	Gender				
	Education level		0	0	
	Father education level		0		
Characteristics of the Environment	Monthly income	0	0	0	0
	Job		0	0	
	Housing types	0			
	Family		0		
Physiological Factors	Chronic diseases number		0	0	
	Chronic diseases total duration		0		
	Hypertension prevalence		0		
	Hypertension diagnosis time		0	0	
	Hyperlipidemia duration		0		
	Hyperlipidemia diagnosis time	0			
	Arthritis doctor diagnose		0		
	Rheumatoid arthritis duration	0	0	0	
	Osteoarthritis duration	0	0	0	
	Osteoarthritis diagnosis time	0	0		
	Osteoporosis prevalence	0	0	0	
	Stroke duration		0		
	Stroke diagnosis time	0	0	0	
	Diabetes duration		0	0	
	Diabetes prevalence		0	0	
	Diabetes diagnosis time		0		
	Tuberculosis duration	0			
	Lung cancer diagnosis time	0			
	Angina diagnosis time		0	0	
	Myocardial infarction, Angina treatment		0	0	0
	Depression doctor diagnose		0		
	Recently 1 month sick days		0	0	0

Factors	Variables	DT	RF	SVM	SLR
	Lumbar T-score	0			
	Femoral neck T-score		0	0	
	Femoral bone T-score		0	0	
	Height variation		0	0	
	Waist circumference	0	0	0	
	Body mass index (BMI)		0	0	
	End-diastolic pressure		0		
	End- systolic pressure		0		
	Insulin (Blood tests)	0	0		
	Fasting blood glucose	0			
	Fasting hours	0	0	0	
	Number of caries permanent teeth	0			
	Number of permanent teeth caries experience	0	0		
	Mandibular prosthesis status		0	0	
Symptom Experience	Last 2 weeks of discomfort days	0	0	0	0
	Last 2 wks of discomfort experience		0	0	
	Continuous depression≥2 weeks	0	0	0	0
	1 year suicidal ideation	0			
	Rhinitis week duration	0	0		
	Restriction of activity: Diabetes		0		
	Restriction of activity: Old-age		0	0	0
	Restriction of activity: Obesity	0	0		
General Health Perceptions	Perceived health status	0	0	0	0
Health Promoting Behaviors	1 week walking days		0	0	
	Vigorous physical activities 1 week		0		
	Moderate physical activity duration		0	0	
	Smokers cigarettes per a day	0	0		
	1 year frequency of drinking		0	0	
	Affect the frequency of daily life due to drinking		0		
	Injury caused by drinking		0		
	Diet reasons: Weight control		0		
	1 years of weight control		0		
	Copayments for outpatient		0	0	
	Outpatient utilization reasons	0	0	0	0
	Pharmacy purchases		0		

4가지 모델에서 모두 중요한 영향요인으로 확인된 변수들은 총 5개로 ‘월 평균 가구총소득’, ‘최근 2주간 불편감 일수’, ‘2주이상 연속 우울감 여부’, ‘주관적 건강상태’, ‘외래이용이유’가 해당되었다.

3가지 모델에서 중요한 영향요인으로 확인된 변수들은 총 9개로 ‘류마티스성 관절염 소요기간’, ‘골관절염 소요기간’, ‘골다공증 유병여부’, ‘뇌졸중 진단시기’, ‘심근경색증, 협심증 치료’, ‘최근 1달간 외병일수’, ‘허리둘레’, ‘공복시간’, ‘활동제한사유(노령)’이 해당되었다.

1~2가지 모델에서 중요한 영향요인으로 확인되었지만, 개별모델에서 우선순위와 가중치가 높았던 변수들로는 ‘만성질환 개수’, ‘만성질환 합산기간’, ‘체질량지수’, ‘고혈압 진단시기’, ‘당뇨병 소요기간’, ‘폐암 진단받은 시기’, ‘대퇴골 전체 T-score’, ‘인슐린(혈액검사)’, ‘하악보철물상태’, ‘활동제한사유(비만)’, ‘1년간 자살생각여부’, ‘비염증상 지속기간’, ‘중등도 신체활동 지속시간’등이 해당되었다.

SLR 적용을 통해 통계적으로 검증된 변수들 외에 ML 적용결과 영향요인으로 도출된 중요변수들의 특성에 대해서 살펴보았으며, 그 결과는 다음과 같다(표 14).

DT에서 중요 영향요인으로 확인된 ‘폐암 진단받은 시기’, ‘비염증상 지속기간’, ‘활동제한사유(비만)’은 비대칭적인 표본의 수에 영향을 받은 결과값으로 판단되어 제외하였다.

Table 14. Mean score of EQ-5D index by influencing factors of ML

Variables		M (SD)	EQ-5D index		X ²	p-value
			Below 0.678	Above 0.678		
Rheumatoid arthritis	0	0.868 (0.150)	65	610	2.92	0.232
duration (yrs)	1~9	0.788 (0.228)	3	17		
	≥10	0.839 (0.082)	0	21		
Osteoarthritis	0	0.891 (0.148)	35	447	13.50	0.001*
duration (yrs)	1~9	0.829 (0.133)	17	137		
	≥10	0.777 (0.160)	16	64		
Osteoporosis prevalence	Normal	0.912 (0.141)	8	156	8.40	0.014*
	Osteopenia	0.868 (0.137)	32	317		
	Osteoporosis	0.821 (0.170)	28	175		
Stroke diagnosis time (age)	0	0.869 (0.150)	59	613	29.36	0.000*
	39~64	0.868 (0.101)	1	26		
	≥65	0.670 (0.123)	8	9		
Diabetes	0	0.864 (0.149)	55	510	0.88	0.645
duration	1~9	0.890 (0.142)	6	81		
(yrs)	≥10	0.838 (0.179)	7	57		
Femur T-score	≥-1.0(Normal)	0.879 (0.145)	40	443	2.56	0.278
	-1.1~-1.24	0.838 (0.155)	25	184		
	≤-2.5	0.812 (0.197)	3	21		
Waist circumference	Normal	0.867 (0.151)	43	387	0.32	0.574
	Obesity	0.861 (0.152)	25	261		
Insulin	2~20μIU/mL	0.863 (0.153)	68	620	3.06	0.080
	≥21μIU/mL	0.894 (0.097)	0	28		
Mandibular prosthesis status	No	0.880 (0.134)	18	235	6.68	0.035*
	Part	0.865 (0.151)	38	356		
	Total	0.807 (0.195)	12	57		
1 year suicidal ideation	Yes	0.768 (0.186)	38	124	47.47	0.000*
	No	0.893 (0.126)	30	524		
Moderate physical activity	0	0.862 (0.157)	46	385	2.98	0.225
duration (min)	10~149	0.869 (0.140)	13	188		
	≥150	0.869 (0.149)	9	75		

* $p < 0.05$ (Fisher's exact test result)

류마티스성 관절염 소요기간에 따른 EQ-5D index 평균값은 10년 이상이 10년 미만에 비해 높았으나, HRQoL 그룹별 차이는 유의하지 않았다.

골관절염 소요기간에 따른 EQ-5D index 평균값은 기간이 증가할수록 낮았으며, HRQoL 그룹별 차이는 유의하였다.

골다공증 유병여부에 따른 EQ-5D index 평균값은 정상 그룹이 가장 높고, 골다공증 그룹이 가장 낮았으며, HRQoL 그룹별 차이는 유의하였다.

뇌졸중 진단시기에 따른 EQ-5D index 평균값은 65세이상 그룹이 가장 낮았으며, HRQoL 그룹별 차이는 유의하였다.

당뇨병 소요기간에 따른 EQ-5D index 평균값은 10년 이상이 10년 미만에 비해 낮았으나, HRQoL 그룹별 차이는 유의하지 않았다.

대퇴골 T-score에 따른 EQ-5D index 평균값은 정상 그룹이 가장 높고, 골다공증 그룹이 가장 낮았으나, HRQoL 그룹별 차이는 유의하지 않았다.

허리둘레와 인슐린(혈액검사)에 따른 EQ-5D index 평균값에 대한 HRQoL 그룹별 차이는 유의하지 않았다.

하악보철물 상태에 따른 EQ-5D index 평균값은 하악보철물이 없는 그룹이 가장 높고, 하악보철물 전체그룹이 가장 낮았으며, HRQoL 그룹별 차이는 유의하였다.

1년간 자살생각 여부에 따른 EQ-5D index 평균값은 예라고 대답한 그룹이 낮았으며, HRQoL 그룹별 차이는 유의하였다. 중등도 신체활동 지속시간에 따른 EQ-5D index 평균값에 대한 HRQoL 그룹별 차이는 유의하지 않았다.

4. 가설모형 수정

연구 결과를 바탕으로 통계적으로 유의성이 검증된 SLR의 결과를 채택하여 가설적 연구모형을 수정하였으며, 최종적인 연구모형은 다음과 같다 (그림 14).

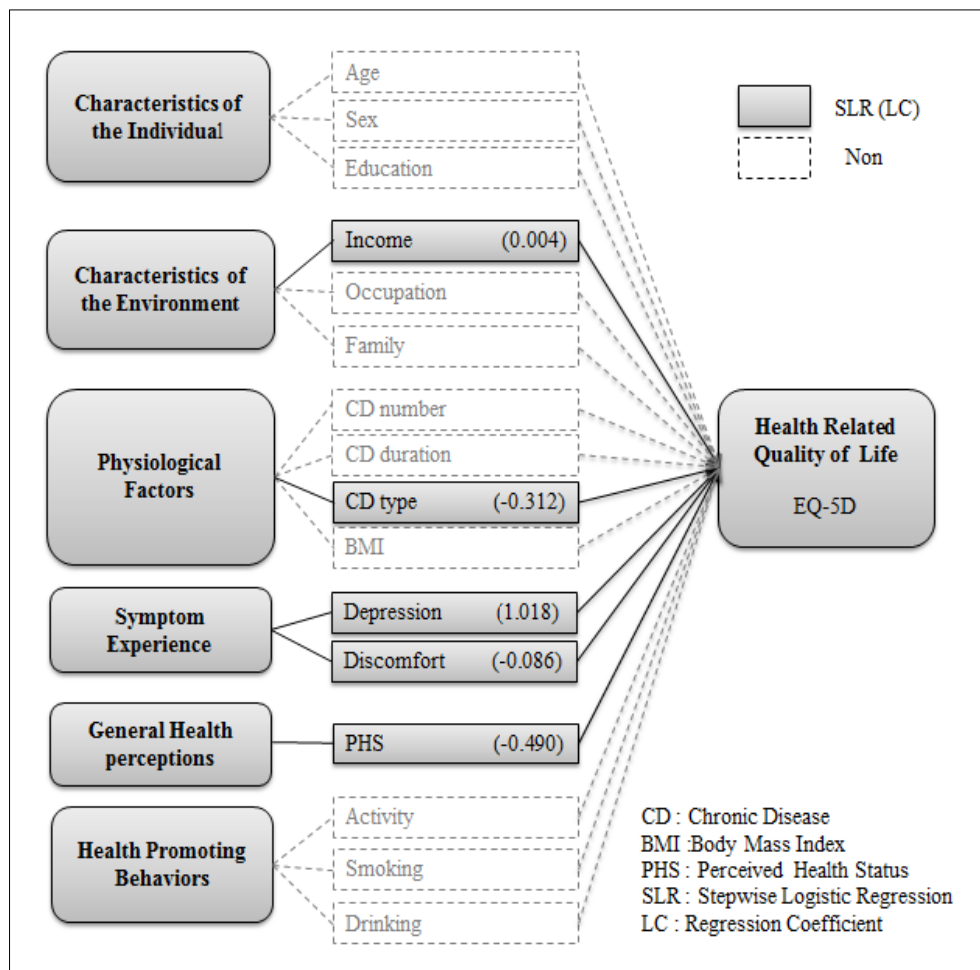


Figure 14. Modified conceptual framework of prediction model of health related quality of life in elderly with chronic diseases

노인 만성질환자의 HRQoL에 영향을 미치는 변수들(소득, 만성질환 종류, 우울, 불편감, 주관적인 건강상태)은 실선으로 회귀계수를 함께 표시하였으며, 영향을 미치지 않는 변수들(연령, 성별, 학력, 직업, 가족, 만성질환 개수, 만성질환 기간, BMI, 활동량, 흡연, 음주)은 점선으로 표시하였다. 본 연구의 노인 만성질환자의 HRQoL 예측모델은 통계적으로 적합성을 증명하였으므로 노인 만성질환자들을 포괄적으로 사정, 중재할 수 있는 이론적 기틀로서 유용하게 사용될 수 있을 것이다.

5. 연구가설 검증

본 연구모형에서 제시한 연구가설의 검증결과를 살펴보면 다음과 같다.

1. ‘연령은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=0.028$, $p=0.866$ 으로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
2. ‘성별은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=0.134$, $p=0.715$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
3. ‘학력은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=2.373$, $p=0.123$ 으로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
4. ‘소득은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=0.004$, $p=0.007$ 로 유의수준 0.05에서 통계적으로 유의하여 지지되었다.
5. ‘직업은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=1.589$, $p=0.208$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
6. ‘가족은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=2.018$, $p=0.155$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
7. ‘만성질환 개수는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=0.427$, $p=0.514$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
8. ‘만성질환 합산기간은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta=1.729$, $p=0.189$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.

9. ‘만성질환 종류는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = -0.312$, $p = 0.038$ 로 유의수준 0.05에서 통계적으로 유의하여 지지되었다.
10. ‘체질량지수는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = 0.891$, $p = 0.345$ 로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
11. ‘우울은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = 1.018$, $p = 0.002$ 로 유의수준 0.05에서 통계적으로 유의하여 지지되었다.
12. ‘불편감은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = -0.0086$, $p = 0.002$ 로 유의수준 0.05에서 통계적으로 유의하여 지지되었다.
13. ‘주관적 건강상태는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = -0.490$, $p = 0.011$ 로 유의수준 0.05에서 통계적으로 유의하여 지지되었다.
14. ‘활동량은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = 0.926$, $p = 0.336$ 으로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
15. ‘흡연은 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = 0.439$, $p = 0.507$ 으로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.
16. ‘음주는 노인 만성질환자의 HRQoL에 영향을 미칠 것이다.’는 SLR을 통해 검증한 결과 $\beta = 0.083$, $p = 0.773$ 으로 유의수준 0.05에서 통계적으로 유의하지 않아 기각되었다.

6. 예측모델 개발

6.1 Decision Tree model

다음은 선정된 변수들을 Input data로 활용하여 MATLAB에서 classificationTree.fit fuction을 통해 DT 모델을 적용하였으며, feature weight을 위해 COST function을 함께 적용하였고, 그 결과는 다음과 같다 (그림 15).

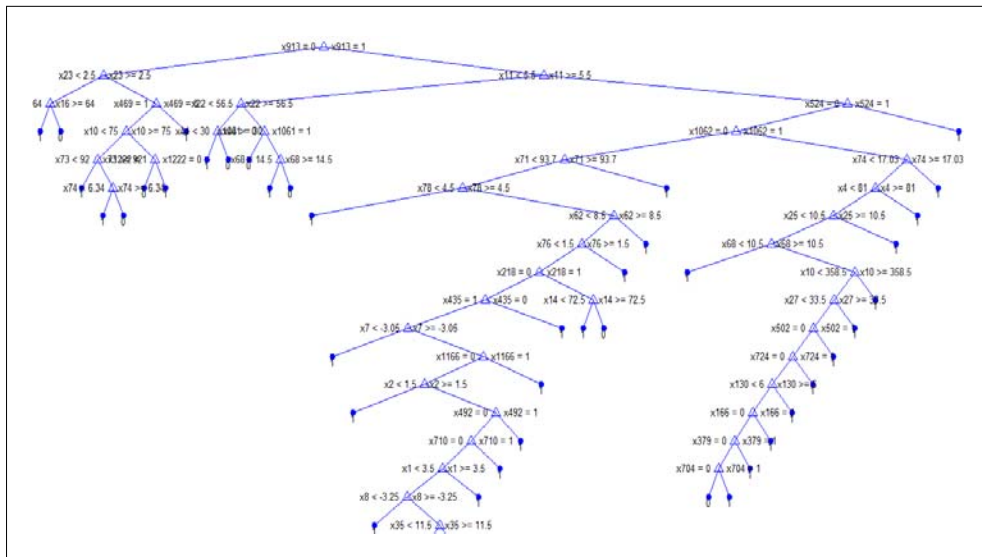


Figure 15. Decision trees graph

‘활동제한사유(비만)’이 아닌 그룹에서 ‘골관절염 소요기간’이 2.5년 미만인 경우는 ‘뇌졸중 진단시기’에 따라 삶의 질 그룹이 결정되었으며, ‘골관절염 소요기간’이 2.5년 이상인 경우는 ‘골다공증 유병여부’에 따라 HRQoL 그룹이 결정되었다. 반면 ‘활동제한사유(비만)’이라고 대답한 그룹은 ‘최근 2주간 불편감 일수’가 5.5일 미만인 경우 ‘골관절염 진단시기’, 5.5일 이상인 경우는 ‘주관적 건강상태’가 중요 요인으로 나타났다.

6.2 Random Forest model

본 연구에서는 예측모델의 Matrix 최적화를 위해 RF기법을 적용하여 Normalized Data로 Feature 선택 및 성능 Baseline을 측정하였다.

전체 독립변수($V^2=1,482$)들을 대상으로 MATLAB에서 'TreeBagger'라는 function을 통해 RF를 적용하였으며, feature weight을 위해 COST function을 함께 적용하였다. Number of trees를 1부터 200까지 증가시키면서 성능을 측정하였으며, 그 결과 4개의 trees일 경우에 F-score가 0.346으로 가장 높게 나타났다(그림 16).

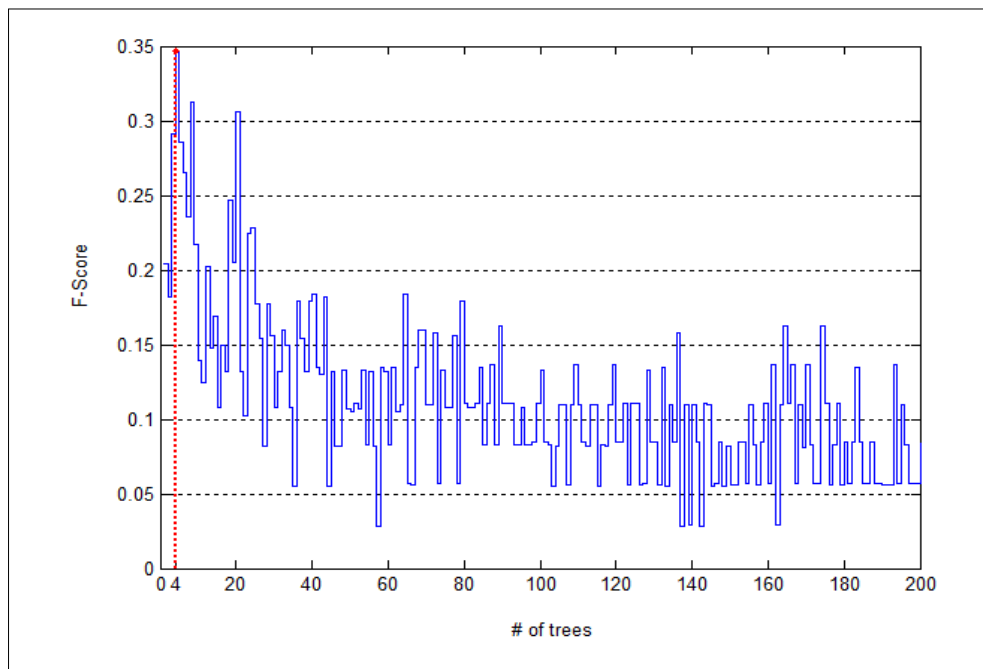


Figure 16. F-score by number of trees

최적의 number of trees를 기준으로 Feature importance를 구하였다 (그림 17).

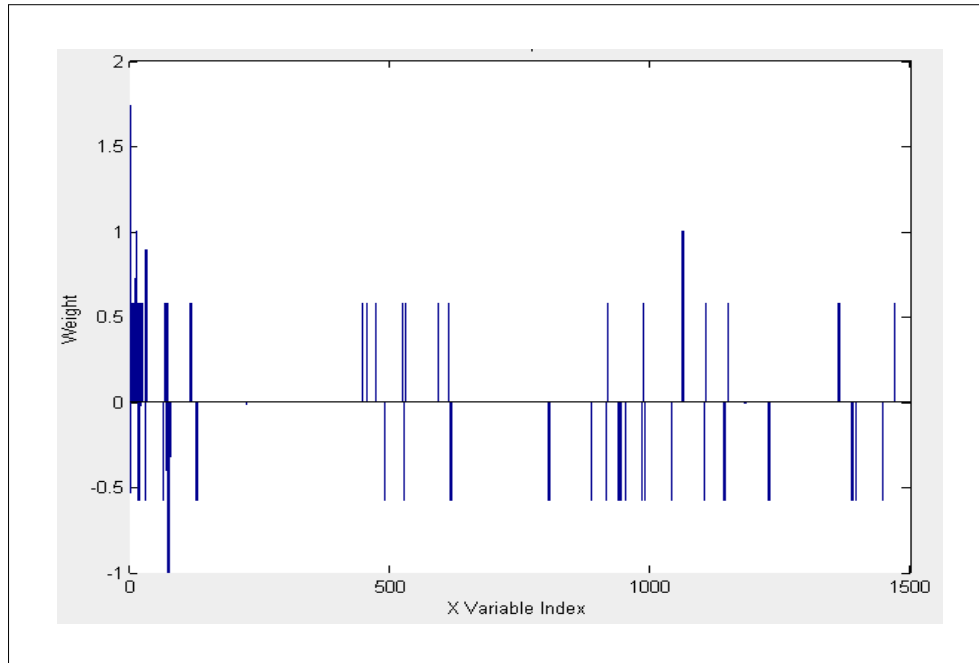


Figure 17. Feature importance

6.3 Support Vector Machine model

MATLAB에서 'svmparam'이라는 function을 통해 SVM을 적용하였으며, 방사형 기저 함수(Radial Basis Function: 이하 RBF) Kernels을 이용한 SVM 분류기를 사용하였다.

우선 RF에서 나온 중요변수 64개를 대상으로 SVM parameter인 c (cost)와 γ (gamma)의 number of features별 성능을 측정하였다. 경험적 오류(empirical error)와 일반화 오류(generalization error)간의 가중치를 나타내는 Parameter c 는 -3부터 3까지 0.2 단위로 증가시켰고, 커널의 매개변수인 Parameter γ 는 -6부터 0까지 0.2단위로 증가시키면서 SVM의 성능을 측정하였다. 그 결과 변수가 50개인 경우 F-score가 0.507로 가장 높게 나타났다(그림 18).

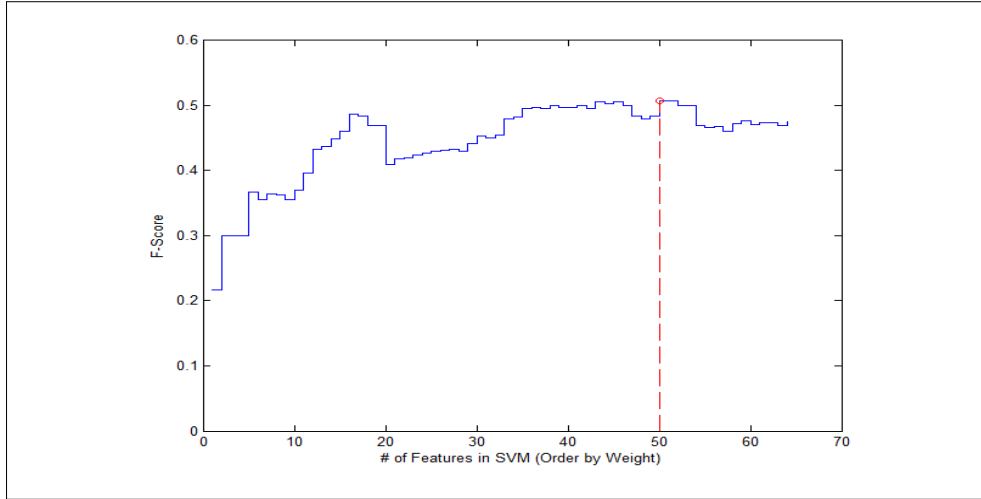


Figure 18. Performance per number of features in SVM

최대 성능을 나타내는 SVM parameter value는 $c=0.819$, $\gamma=0.3012$ 가 해당된다. Sample 수의 보정을 위해 Feature weight은 10을 지정하였으며, 다른 parameter value인 c 와 γ 를 증가시키면서 측정한 SVM performance에 대한 결과는 다음과 같다(그림 19).

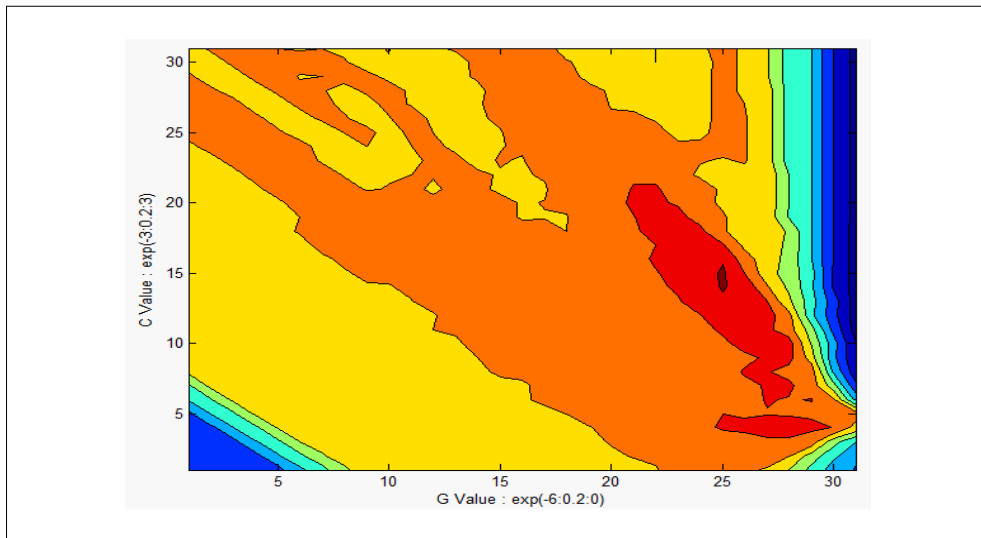


Figure 19. Performance contour with parameter c & γ in SVM

6.4 Stepwise Logistic Regression model

로지스틱 회귀분석은 독립적인 여러 자질들을 입력으로 사용하여 이항확률을 계산하기 위해 사용되는 일반화된 선형모델로 본 연구에서는 SPSS에서 단계적 로지스틱 회귀분석을 적용하였으며, 유의수준은 $p < 0.05$ 로 통계적 유의성을 검증하였다.

단계적 변수 선택은 8단계에 거쳐 종료되었으며, 회귀모형은 유의한 것으로 나타났다($p=0.000$). 모형의 설명력을 나타내는 Cox & Snell의 결정계수(R^2)에 의하면 20.0%의 설명력을 보여주고 있으며, Nagelkerke의 결정계수(R^2)도 42.9%의 설명력을 보여주고 있다.

총 8개의 변수가 유의하게 나왔으며, ‘2주이상 연속 우울감 여부($p=0.002$)’, ‘심근경색증, 협심증 치료($p=0.038$)’, ‘외래 이용이유($p=0.004$)’, ‘월평균 가구총소득($p=0.007$)’, ‘주관적 건강상태($p=0.011$)’, ‘최근 1달간 와병일수($p=0.007$)’, ‘최근 2주간 불편감일수($p=0.002$)’, ‘활동제한사유(노령)($p=0.002$)’이 해당되었다.

7. 예측모델 평가

7.1 Decision Tree model

DT 모델의 10 fold cross-validation에 대한 분류행렬(Confusion matrix: 이하 CM)은 다음과 같다(표 15). CM은 전형적으로 시각적인 툴로서 사용되며, 행렬에 대한 행은 모델의 예측값을 나타내고 열은 실제값을 나타낸다.

Table 15. Confusion matrix of DT

Actual class	Predicted class	
	0 (Below 0.678)	1 (Above 0.678)
0 (Below 0.678)	True Positive	False Negative
1 (Above 0.678)	False Positive	True Negative
1	2	4
	8	57
2	4	3
	9	56
3	2	5
	3	62
4	0	7
	9	56
5	3	4
	9	56
6	0	7
	12	53
7	1	6
	7	58
8	1	6
	6	58
9	1	6
	7	57
10	1	5
	6	59

DT 모델의 분류예측에 대한 평균 Performance를 나타내는 Box plot 도표는 다음과 같다(그림 20). Accuracy는 0.82, F-score는 0.23을 나타내었다.

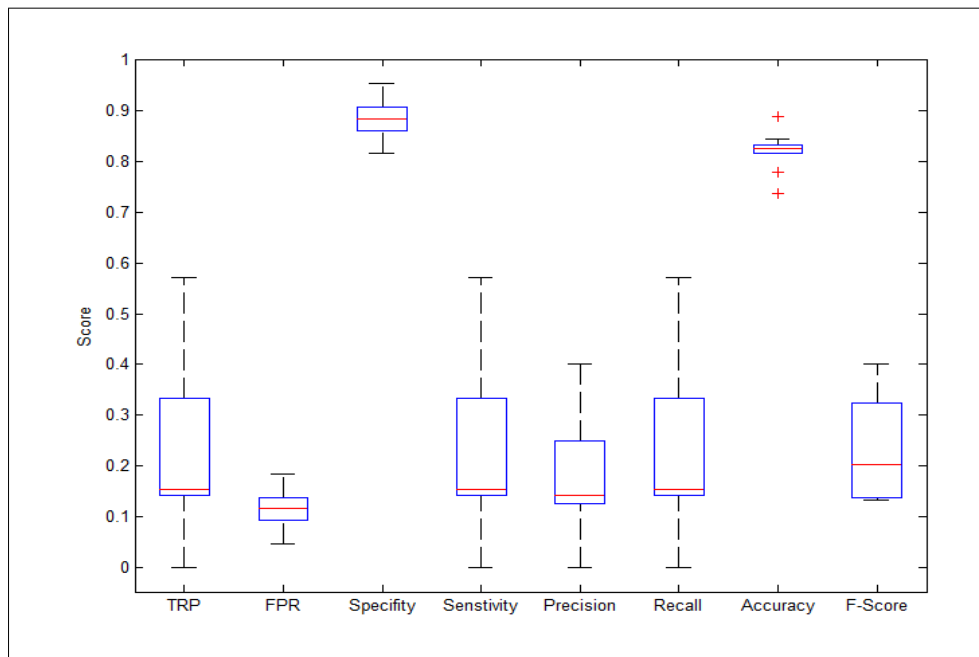


Figure 20. Boxplot of performance in DT

7.2 Random Forest model

RF 모델의 10 fold cross-validation에 대한 CM은 표 16과 같다.

Table 16. Confusion matrix of RF

Actual class	Predicted class	
	0 (Below 0.678)	1 (Above 0.678)
0 (Below 0.678)	True Positive	False Negative
1 (Above 0.678)	False Positive	True Negative
1	1 2	5 63
2	3 8	4 57
3	2 4	5 61
4	4 4	3 61
5	2 6	5 59
6	3 3	4 62
7	0 4	7 61
8	2 5	5 59
9	2 3	5 61
10	1 4	5 61

RF 모델의 분류예측에 대한 평균 Performance를 나타내는 Box plot 도표는 다음과 같다(그림 21). Accuracy는 0.87, F-score는 0.33를 나타내었다.

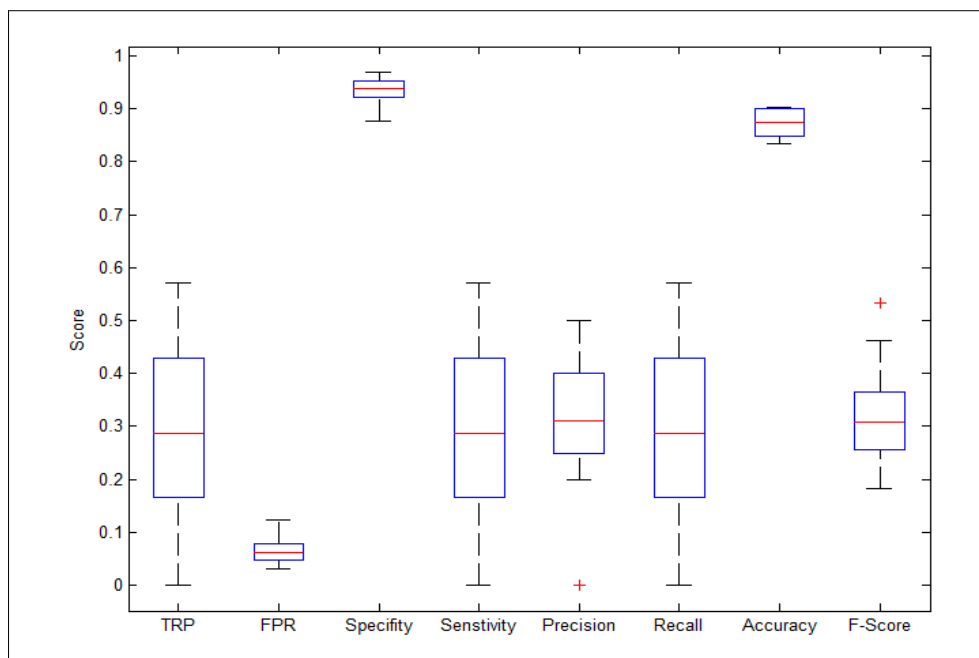


Figure 21. Boxplot of performance in RF

7.3. Support Vector Machine model

SVM 모델의 10 fold cross-validation에 대한 CM은 표 17과 같다.

Table 17. Confusion matrix of SVM

Actual class	Predicted class	
	0 (Below 0.678)	1 (Above 0.678)
0 (Below 0.678)	True Positive	False Negative
1 (Above 0.678)	False Positive	True Negative
1	2 1	4 64
2	4 2	3 63
3	3 4	4 61
4	4 5	3 60
5	3 4	4 58
6	4 0	3 65
7	3 7	4 58
8	6 3	1 61
9	3 5	4 59
10	4 6	2 59

SVM 모델의 분류예측에 대한 평균 Performance를 나타내는 Box plot 도표는 다음과 같다(그림 22). Accuracy는 0.90, F-score는 0.51을 나타내었다.

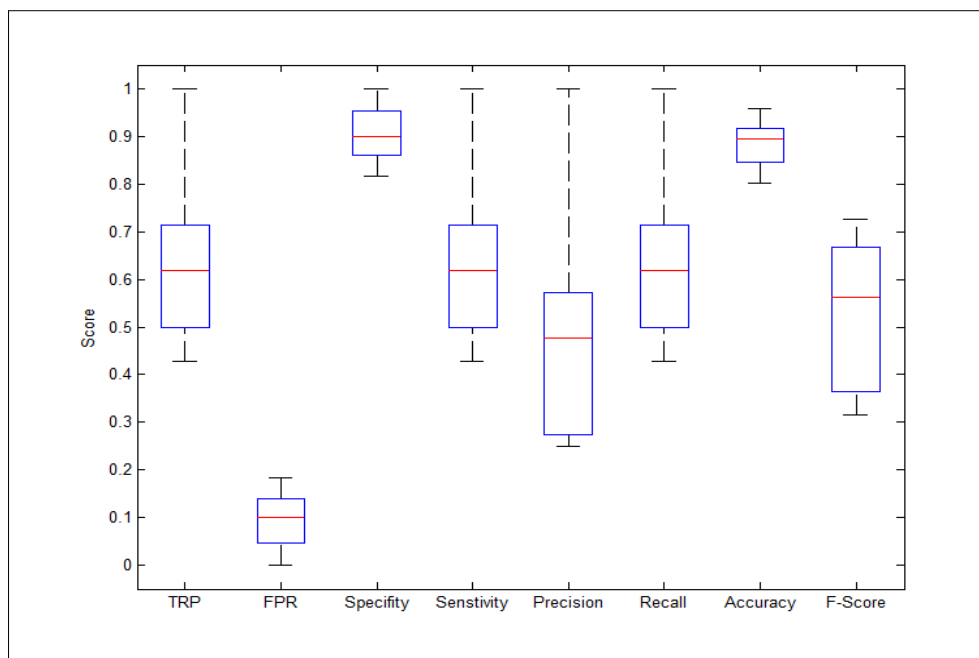


Figure 22. Boxplot of performance in SVM

7.4. Stepwise Logistic Regression model

SLR 모델의 10 fold cross-validation에 대한 CM은 표 18과 같다.

Table 18. Confusion matrix of SLR

Actual class	Predicted class	
	0 (Below 0.678)	1 (Above 0.678)
0 (Below 0.678)	True Positive	False Negative
1 (Above 0.678)	False Positive	True Negative
1	20 8	38 578
2	24 9	38 573
3	19 9	43 573
4	23 6	36 579
5	22 9	41 572
6	22 6	36 580
7	25 8	35 576
8	23 8	39 574
9	24 8	40 572
10	23 7	37 577

SLR 모델의 분류예측에 대한 평균 Performance를 나타내는 Box plot 도표는 다음과 같다(그림 23). Accuracy는 0.93, F-score는 0.49를 나타내었다.

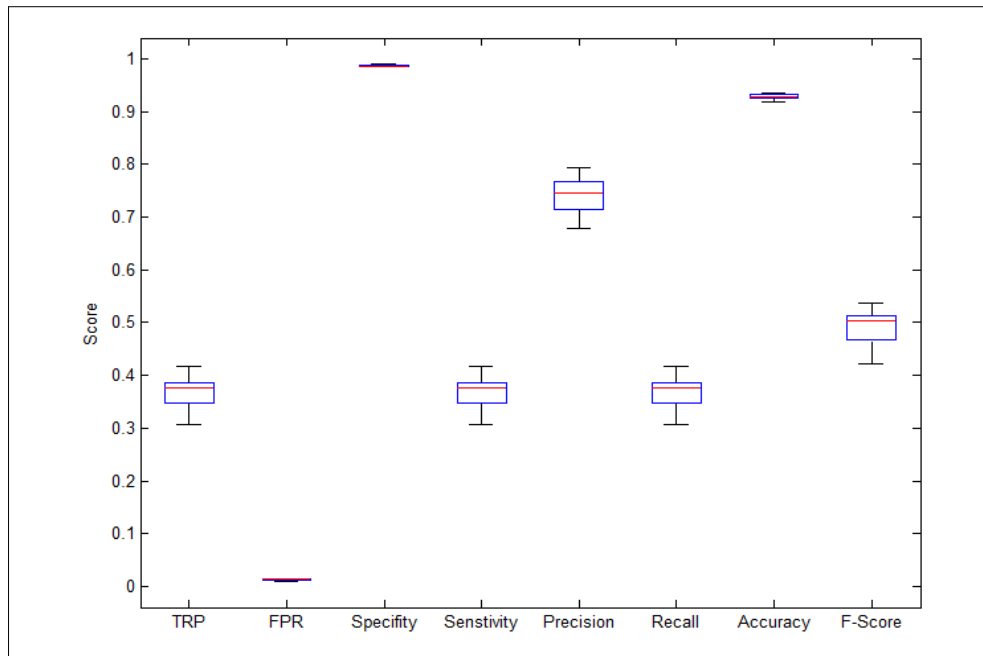


Figure 23. Boxplot of performance in SLR

7.5 Comparison of all prediction models

노인 만성질환자의 HRQoL 위험군 예측을 위한 RF, DT, SVM, SLR 모델의 성능을 각각 비교 평가하면 다음과 같다(표 19).

Table 19. Comparison of performance in prediction models

Name	DT	RF	SVM	SLR
TP Rate	0.22	0.29	0.53	0.37
FP Rate	0.12	0.07	0.06	0.01
Specificity	0.88	0.93	0.94	0.99
Sensitivity	0.22	0.29	0.53	0.37
Precision	0.17	0.31	0.53	0.74
Recall	0.22	0.29	0.53	0.37
Accuracy	0.82	0.87	0.90	0.93
F-score	0.23	0.33	0.51	0.49

본 연구에서는 4가지 모델 중에서 SLR이 Accuracy=0.93, F-score=0.49로 가장 우수한 성능을 나타내었으며, 그 다음으로는 SVM(0.90, 0.51), RF(0.87, 0.33), DT(0.82, 0.23) 순이었다. 4가지 모델의 성능 지표별 비교 결과를 방사형 차트로 살펴보면 다음과 같다(그림 24).

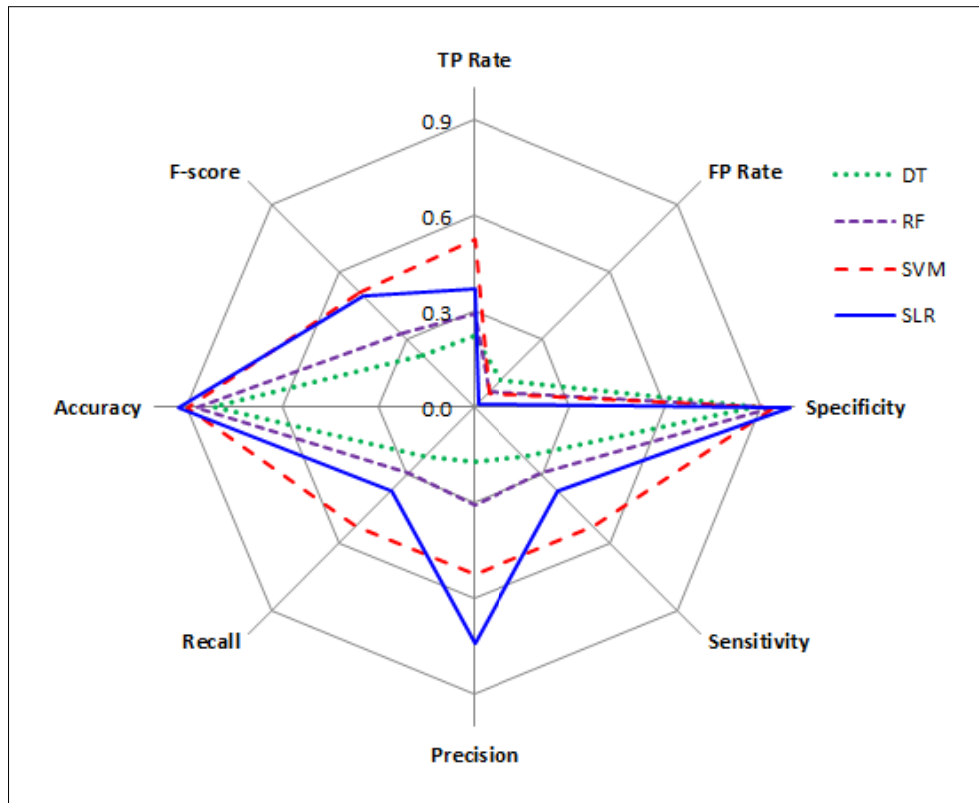


Figure 24. Rader chart of performance of four models

본 연구에서의 예측모델에 대한 평가는 다음의 2가지 측면을 모두 포함하여 판단하였다. 우선 평가지표를 비교하였을 때 SLR 모델이 가장 좋은 성능을 나타내었다. 또한 예측모델 개발에 사용된 변수 구성의 수를 비교하였을 때 DT는 1,482개, RF는 64개, SVM ($c=0.819$, $\gamma=0.3012$, $w=10$)은 32개를 포함하기 때문에 가장 적은 8개의 변수를 포함하여 예측한 SLR 모델이 본 연구에서 노인 만성질환자의 HRQoL 그룹을 예측할 수 있는 가장 최적의 모델이라고 할 수 있다. 모델 개발 시 사용했던 MATLAB code를 부록 5로 첨부하였다.

8. 기계학습 기법의 적용가능성 탐색

본 연구에서 노인 만성질환자의 HRQoL 위험군 예측모델 개발에 있어 기계학습 기법을 적용한 결과 SVM의 경우 SLR보다 약간 저하된 성능을 나타내었으며, SLR에 비해 사용된 변수들이 많아 예측모델의 효율성 측면에서도 개선이 요구되었다. 따라서 성능 및 효율과 관련된 문제점들에 대한 해결방안을 살펴볼 필요가 있기에 다음과 같이 자료 정규화에 대한 부분적인 실험을 실시하였다(표 20).

Table 20. Re-examination sets

Criteria	File	Variables	Features	Binary	Continuous variables	Ordinal variables	Normal variables
Original	V22	498	1,482	O(A)	134	25	339
	V23	498	498	X	134	25	339
SLR	V24	8	20	O	3	-	5
	V25	8	20	O	-	3	5
	V26	8	16	O	-	4	4
	V27	8	27	O	-	-	8
HC1 (Discrete set)	V28	11	84	O	2	-	9
	V29	11	84	O	-	2	9
HC2 (Continuous set)	V30	11	11	X	11	-	-
	V31	11	11	X	-	11	-
	V32	11	37	O	-	-	11

SLR에서 유의하다고 확인된 총 8개의 변수들과 계층적 군집(Hierarchical Clustering: 이하 HC) 분석을 통해 선택한 총 11개의 변수로 세트를 구성하였다(표 21). 자료 정규화를 통해 SVM 모형의 예측 성능을 비교해보았다.

Table 21. Variables of re-examination sets

Set	Variables	
SLR	Continuous variables	<ul style="list-style-type: none"> ▪ Monthly income ▪ Last 2 weeks of discomfort days ▪ Recently 1 month sick days
	Discrete variables	<ul style="list-style-type: none"> ▪ Perceived health status ▪ Myocardial infarction, angina treatment ▪ Restriction of activity: Old-age ▪ Continuous depression ≥ 2 weeks ▪ Outpatient utilization reasons*
HC1 (Discrete variables)	Continuous variables	<ul style="list-style-type: none"> ▪ Chronic diseases number ▪ Chronic diseases total duration
	Discrete variables	<ul style="list-style-type: none"> ▪ Outpatient utilization reasons* ▪ Education level ▪ Perceived health status ▪ Last 2 wks of discomfort experience ▪ Myocardial infarction, angina treatment ▪ Arthritis doctor diagnose ▪ Continuous depression ≥ 2 weeks ▪ Diabetes prevalence ▪ Perceived oral health status
HC2 (Continuous variables)	Continuous variables	<ul style="list-style-type: none"> ▪ Chronic diseases number ▪ Chronic diseases total duration ▪ Age ▪ Femoral bone T-score ▪ Last 2 weeks of discomfort days ▪ Recently 1 month sick days ▪ Osteoarthritis duration ▪ Rheumatoid arthritis duration ▪ Diabetes duration ▪ Body mass index ▪ Copayments for outpatient

* Outpatient utilization reasons : Disease, Accident & poisoning, Others

HC 분석을 통해 다음과 같은 결과(그림 25)를 얻었으며, 그룹별로 대표적인 변수를 전문가 지식에 의해 선별하였다.

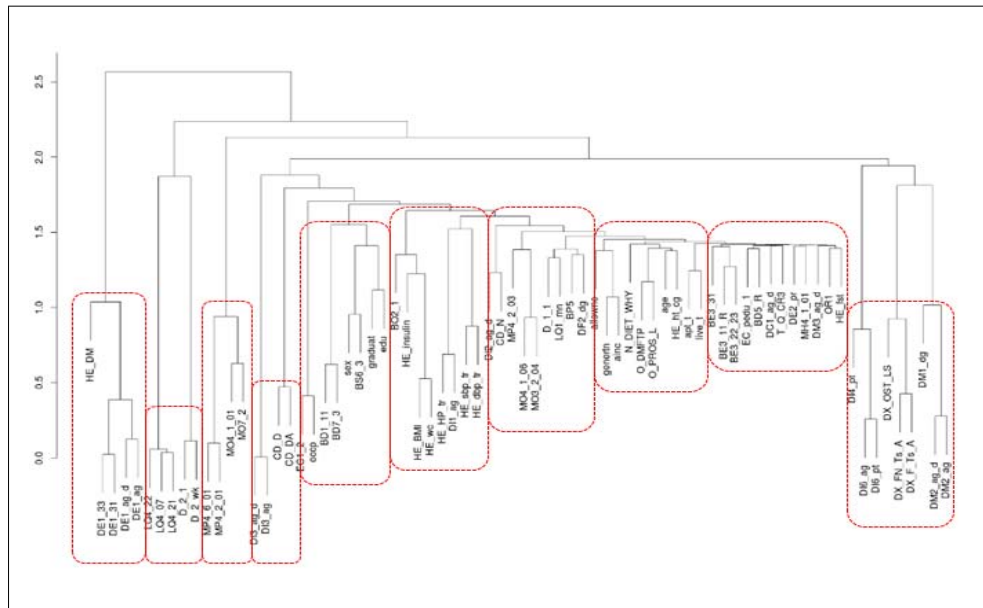


Figure 25. Dendrogram of hierarchical clustering

‘당뇨병 유병여부, 당뇨병 치료, 당뇨병 진단시기’ 그룹과 ‘활동제한사유, 최근 2주간 몸이 불편했던 경험, 최근 2주간 불편감 일수’ 그룹, ‘뇌졸중 소요기간, 뇌졸중 진단시기, 만성질환 흡산기간, 만성질환 개별기간’ 그룹 등으로 구분되었다.

구분된 그룹들을 참조하여 다음과 같이 이산형 변수중심의 HC1 set와 연속형 변수중심의 HC2 set로 구성하였으며, 표 22와 같이 변수들을 정규화하였다.

Table 22. Variables of hierarchical clustering set

No	HC1 variables	Range	HC2 variables (M)	Discretization
1	Chronic diseases number (ea)	1~8 (Mean 1.95)	Chronic diseases number (ea)	1. 1 2. 2 3. 3 4. ≥ 4
2	Chronic diseases total duration (yrs)	0~174 (Mean 20)	Chronic diseases total duration (yrs)	1. ~5 2. 6~14 3. 15~30 4. ≥ 31
3	Outpatient utilization reasons	1~299 (Disease code)	Age (Mean 70.6)	1. 65 2. 66~70 3. 71~75 4. 76~89
4	Education level 1. Below elementary 2. Middle school graduate 3. High school graduate 4. Above college graduate		Femoral bone T-score (Mean -0.60)	1. -3.2~-1.3 2. -1.2~-0.6 3. -0.5~0.0 4. 0.1~2.7
5	Perceived health status	1. Very good 2. Good 3. Fair 4. Poor 5. Very poor	Last 2 weeks of discomfort days (Mean 4.4)	1. 0 2. 1~7 3. 8~14
6	Last 2 wks of discomfort experience	1. Yes 2. No	Osteoarthritis duration (Mean 2.7 yrs)	1. 0 2. 1~6 3. 7~34
7	Myocardial infarction, angina treatment	0. No 1. Yes 8. Not applicable	Rheumatoid arthritis duration (Mean 0.7 yrs)	1. 0 2. 1~49
8	Arthritis doctor diagnose	0. No 1. Yes 8. Not applicable	Diabetes duration (Mean 2.1 yrs)	1. 0 2. 1 3. 2~36
9	Continuous depression ≥ 2 weeks	1. Yes 2. No	Body mass index (Mean 23.9)	1. 15.6~21.8 2. 21.9~23.8 3. 23.9~25.9 4. 26.0~40.5
10	Diabetes prevalence	1. Normal 2. Impaired 3. Diabetes	Copayments for outpatient (Mean 10,189 won)	1. 0 2. 1~1,500 3. 1,501~2,700 thousand won
11	Perceived oral health status	1. Very good 2. Good 3. Fair 4. Poor 5. Very poor	Recently 1 month sick days (Mean 0.7)	1. 0 2. 1~2 3. 3~30

부분 실험의 변수 세트 선정과 자료 정규화 과정을 통하여 SVM 을 적용한 결과 다음과 같은 최종 결과를 얻었다(표 23).

Table 23. Result of re-examination

Criteria	File	Binary	Cont	Ordi	Norm	Accuracy	F-score
Original 1	V22	O(A)	134	25	339	0.90	0.51
HC2 (Cont)	V32	O	-	-	11	0.87	0.31
HC1 (Disc)	V29	O	-	2	9	0.87	0.40
HC1 (Disc)	V28	O	2	-	9	0.85	0.39
Original 2	V23	X	134	25	339	0.84	0.42
SLR	V25	O	-	3	5	0.82	0.43
SLR	V24	O	3	-	5	0.81	0.43
SLR	V27	O	-	-	8	0.81	0.40
SLR	V26	O	-	4	4	0.78	0.40
HC2 (Cont)	V30	X	11	-	-	0.72	0.34
HC2 (Cont)	V31	X	-	11	-	0.69	0.29

자료 정규화 과정을 통해 실험군을 변형하였으며, 부분 실험의 경우 원래 SVM 모델의 Parameter 값을 고정한 상태로 적용하였기 때문에 부분 실험의 결과가 원래 모형의 성능을 능가하지 않았으나 실험군간의 상대적인 성능 비교는 가능하다고 보았다. 변수의 이산화는 동일크기 백분위수로 구간화하는 방법을 적용하였다.

모델에서 사용했던 원자료인 V22가 이산화하지 않은 자료들로 변형한 V23보다 예측 성능이 높았다. 따라서 자료를 1-of-K encoding 하는 방법이 SVM의 성능을 높인다는 것을 알 수 있었다.

SLR Set로 구성된 V25가 V24보다 성능이 높았으며, V27이 V26보다 성능이 높게 나왔다. 이는 자료를 이산화하여 구간변수로 변형할수록 SVM의

성능이 높아진다는 것을 알 수 있었다. 그러나 구간변수가 더 적은 V25가 전체가 구간변수인 V27보다 성능이 높게 나온 것은 보다 적절한 이산화 방법을 적용할 필요가 있음을 시사한다.

HC 세트로 구성된 V32, V29, V28, V30, V31의 순으로 성능이 높은 결과를 나타낸 것은 변수를 구간변수로 이산화할수록 SVM 성능이 높은 것을 알 수 있었으나, 전체가 연속변수인 V30이 전체가 순서변수인 V31보다 성능이 높게 나온 것은 보다 적절한 이산화 방법을 적용할 필요가 있음을 시사한다.

기존 SVM 모델은 32개 변수를 투입하였을 때 최대 90%의 예측 성능을 나타낸 반면에 재실험 모델에서는 11개의 변수만을 투입하여 최대 87%의 예측 성능을 나타내었다. 이는 결론적으로 더 적은 변수를 투입하여 그와 유사한 예측 성능을 나타낼 수 있는 효율적인 SVM 모델의 가능성을 제시하는 것이다.

본 연구는 노인 만성질환자의 HRQoL 예측모델 개발에 있어서 기계학습 기법의 적용에 따르는 성능과 효율성의 문제를 변수 정규화 과정과 변수 선택 과정을 통해서 해결할 수 있음을 부분 실험을 통해 증명하였다. 또한 혼합형태 자료의 ML 기법 적용에 대한 구체적인 가이드라인으로서 제시된 변수 선택 및 자료 정규화 방법표(부록 4)를 참조하여 향후 유사 자료의 분석 및 예측에 활용할 수 있을 것이다.

VI. 논의

인구 고령화와 이로 인한 만성질환의 증가는 결과적으로 노인들의 HRQoL을 저하시키는 원인이 되고 있다. 따라서 노인 만성질환자의 HRQoL 향상을 위해서는 영향요인과 중재가 필요한 위험군에 보다 집중해서 효율적인 건강관리 전략 및 중재방안을 제시할 필요가 있다. 본 연구는 노인 만성질환자들의 HRQoL에 대한 이해를 바탕으로 이를 증진시킬 수 있는 구체적인 실천방안을 개발하는데 도움이 될 것이다. 또한 본 연구에서는 노인 만성질환자들의 HRQoL 분석을 위해 다양한 형태의 수많은 자료들이 생성되는 빅데이터 시대에 적합한 지능형 분석기술로서 기계학습 기법을 최초로 적용하여 그 가능성을 탐색하였다는 점에서 의의가 있다고 할 수 있다.

1. 연구결과에 대한 논의

본 연구에서 노인 만성질환자의 HRQoL에 영향을 미치는 요인들로 ‘소득’, ‘만성질환 종류’, ‘우울’, ‘불편감’, ‘주관적인 건강상태’가 해당되었다. 즉, 소득 수준에 따라, 만성질환 종류에 따라, 2주 이상 연속 우울감 여부에 따라, 최근 2주간 불편감 일수에 따라, 주관적인 건강상태에 따라 노인 만성질환자의 HRQoL이 달라질 수 있음을 알 수 있다. 반면 연령, 성별, 학력, 직업, 가족, 만성질환 개수, 만성질환 기간, BMI, 활동량, 흡연, 음주에 따라서는 노인 만성질환자의 HRQoL에 유의한 차이를 보이지 않았다. 따라서 이러한 결과를 토대로 노인 만성질환자의 HRQoL 향상을 위해서는 이러한 영향요인에 보다 집중해서 건강관리 전략 및 중재방안을 제시할 필요가 있다.

우선 본 연구에서는 소득수준에 따라 노인 만성질환자의 HRQoL이 달라진다는 기존의 연구결과(성상석, 2004; 정영호, 2011; 남상권, 2011)와 동일하게 소득수준이 증가할수록 HRQoL이 증가되는 것을 볼 수 있었다. 노

인들의 경우 경제적인 문제로 자신들의 건강관리 또는 질병치료를 유보할 경향이 높다. 따라서 노인 만성질환자의 HRQoL 향상을 위해서 기초수급대상자와 같은 취약그룹을 우선으로 만성질환 관리를 위한 사회적 혜택 및 경제적인 지원이 요구된다.

남상권(2011)은 만성질환 유형별로 만성질환 노인의 HRQoL 영향요인을 연구한 결과 전체적으로는 연령, 성별, 학력, 가족이 영향을 미쳤으나, 만성질환 유형별로는 부분적으로만 영향을 미치는 것으로 보고하였다. 즉, 내분비계 질환인 경우는 연령, 소득은 영향을 미쳤으나, 근골격계 질환인 경우는 연령, 성별, 학력, 가족이 모두 영향을 미치지 않았다. 따라서 만성질환 특성에 따라서 영향요인이 달라질 수 있음을 고려한다면 본 연구와 같은 결과를 나타낼 수 있을 것이다. 또한 기존 연구에서는 HRQoL을 연속변수로서 영향요인을 파악한 반면, 본 연구에서는 HRQoL을 이분화하여 영향요인을 파악한 점에서도 차이를 나타낼 수 있다고 판단된다. 따라서 HRQoL 향상을 위해서 만성질환의 유형별로 서로 다른 중재방안을 모색해야 할 것이다.

성인의 경우 직업이 있을수록, 직업의 종류에 따라서는 관리직, 전문직 등이 농업직, 서비스직보다 HRQoL 수준이 더 높은 것으로 나타났다(정영호, 2011). 또한 노인의 경우에도 직업 유무에 따라서 HRQoL이 달라진다고 하였다(이동호, 2011). 이처럼 선행연구에서 성인 또는 노인 전체를 대상으로 하였을 경우에는 직업 유무와 종류가 HRQoL에 영향을 미쳤지만, 본 연구와 같이 만성질환을 가진 노인으로 대상을 국한하였을 때는 직업이 영향을 미치지 않는 것으로 나타났다. 따라서 일반 노인의 HRQoL을 향상시키기 위해서는 노인 일자리 찾기 등의 취업에 관련된 중재가 필요하겠지만 만성질환을 가진 노인의 경우는 취업보다는 만성질환에 잘 적응할 수 있도록 돕는 차별화된 접근방법이 요구된다.

질환의 종류나 기간 등이 동일하다 하더라도 대상자마다 통증의 정도나 일상활동 정도, 심리적인 상태가 다르다고 보고되고 있다(Maruta, Vatterott, & McHardy, 1989). 노인 만성질환자에게 필요한 태도로는 만성질환의 수용, 자신감의 함양, 자기관리 수행정도에 따라 건강상태 등이

달라지며 궁극적으로는 삶의 질이 좌우된다고 하였다(송미순, 2004).

본 연구의 단변량 분석에 의하면 만성질환 개수와 만성질환 합산기간이 증가할수록 무조건 HRQoL이 저하되는 것이 아니라 어떤 시점을 기준으로 다시 상승하는 경향을 보이는데 이는 만성질환에 적응함으로써 대응능력이 더 강화되었기 때문이라고 볼 수 있다(김명자 & 송경애, 1991). 실제로 5개 이상의 만성질환을 가진 노인들이 4개의 만성질환을 가진 노인들보다 HRQoL이 높다는 사실은 만성질환의 적응과 대응기전에 대한 구체적인 연구가 HRQoL 연구와 함께 진행되어야 함을 시사한다.

성인의 경우 과체중 및 비만군에서 낮은 HRQoL 수준을 나타내었으며(신상진, 2010), 지역사회 노인의 경우 비만인 노인이 정상체중인 노인에 비해 HRQoL이 낮았다(Groessler, Kaplan, Barrett-Connor, & Ganiats, 2004). 이는 본 연구 결과와 일치하지 않았으나, 체질량지수가 HRQoL에 유의한 관련을 보이지 않은 Park, Shin, and Kim (2000)의 연구결과와는 일치하였다. HRQoL과 체질량지수에 관해서는 아직까지 논란이 있으므로, 이러한 결과의 차이가 어떤 요인에서 기인하는지를 명확하게 밝히는 후속연구가 필요할 것이다.

본 연구에서는 우울이 HRQoL의 영향요인으로 나타났으며, 이는 노인을 대상으로 한 선행연구의 결과와 동일하였다(손신영, 2006; 최윤진, 2009). 질병으로 인해 신체적 불편감이 만성화되고, 이로 인해 HRQoL에 부정적인 영향을 끼친다고 보고한 Holmes (1985)의 연구는 불편감이 HRQoL의 영향요인으로 확인된 본 연구와 동일한 결과를 나타내었다. 또한 주관적 건강상태가 HRQoL에 중요한 영향요인이라고 보고한 많은 선행연구(최수정, 2001; 신상진, 2010; 김정연, 2010; 이동호, 2011)들과 본 연구의 결과가 일치하였다.

노인 만성질환자의 경우 만성질환을 가지고 있지 않는 노인에 비해 HRQoL이 저하된 것은 사실이나, 단순히 만성질환과 관련된 신체적인 측면에 국한해서 접근하기 보다는 우울과 불편감을 완화시키고, 주관적 건강상태를 증진시킬 수 있는 방안을 검토하여 포괄적으로 접근하는 것이 HRQoL

향상시키기 위한 효율적인 중재방안이 될 수 있을 것이다.

노인들의 건강증진 행위(활동량, 흡연, 음주)와 HRQoL 간에 유의한 상관관계가 있으며, 당뇨병 노인의 경우 음주량, 걷기 지속시간이 HRQoL에 영향을 미친다고 보고한 선행연구(최정숙, 2004; 신지원, 2011)와 달리 본 연구에서는 노인 만성질환자들의 건강증진 행위에 따라 HRQoL이 유의한 차이를 나타내지 않았다. 반면 노인의 경우 흡연, 음주, 운동 등의 건강행위가 HRQoL에 직접적인 영향을 미치지 않는다는 김정연(2010)의 연구와 농촌 노인을 대상으로 한 손신영(2006)의 연구와 일부 지역사회 노인을 대상으로 흡연과 음주 및 운동여부 등의 생활습관 관련 특성은 HRQoL과 관련이 없는 것으로 보고한 J. Lim, Park, Kang, and Ryu (2007)의 연구와는 일치하였다.

만성질환을 가지고 있지 않는 노인의 경우 만성질환을 예방하는 건강증진 실천행위를 통하여 HRQoL을 증진시킬 수 있겠으나 이미 만성질환을 가지고 있는 노인들을 대상으로는 건강증진 실천행위에 대하여 차별화된 접근법이 필요하기에 이를 규명하기 위한 더욱 체계적인 연구가 필요할 것으로 생각된다.

본 연구에서는 전통적인 통계방법인 SLR과 상호보완적으로 ML 기법을 적용한 결과, 노인 만성질환자의 HRQoL 예측을 위한 중요변수로서 ‘골관절염 소요기간’, ‘골다공증 유병여부’, ‘뇌졸중 진단시기’, ‘하악보철물 상태’, ‘1년간 자살생각여부’ 등이 추가적으로 도출되었다. 이러한 요인들은 HRQoL 예측에 영향을 미치는 기여변수들이나 통계적으로 검증되지 않았기 때문에 본 연구의 가설모형에서는 제외하였다.

ML 적용결과 확인된 중요한 영향요인들을 중심으로 단변량 및 다변량 분석을 시행하였다. 이들 요인들이 다변량 분석에서는 노인 만성질환자의 HRQoL과 유의한 결과를 나타내지 않았지만, 단변량 분석에서는 ‘골관절염 소요기간’, ‘골다공증 유병여부’, ‘뇌졸중 진단시기’, ‘1년간 자살생각 여부’, ‘하악보철물 상태’가 유의하게 그룹별로 차이가 있음을 알 수 있었다. 즉, 노인 만성질환자의 경우 골관절염 소요기간이 길고, 골다공증이 있으며, 뇌

졸중 진단시기가 늦고, 1년간 자살생각을 한 적이 있으며, 전체적인 하악보철물을 지니고 있을수록 HRQoL 수준이 낮았다.

무릎 골관절염 환자의 HRQoL에 영향을 미치는 요인으로 통증, 우울, 자기효능감이 해당되었으며, 골관절염 소요기간에 따른 그룹간 차이가 있었으나 영향요인에는 해당되지 않았다(김혜민, 2012). 이는 본 연구결과와 동일하였다.

관절염, 골다공증, 디스크 등의 관절질환은 운동능력, 일상활동, 통증/불편의 문제와 관련이 있고, 뇌졸중은 자기관리의 문제와 관련이 있으나 고혈압과 당뇨는 유의미한 관련이 없어 질환의 종류에 따라 HRQoL이 다르게 나타났다(Kang et al., 2007). 이는 골다공증 여부에 따라 HRQoL의 차이를 나타내지 않은 본 연구의 결과와 상이함을 알 수 있다. 따라서 향후 개별 질환 및 복합질환과 노인 만성질환자의 HRQoL과의 관련성에 대한 후속연구를 통하여 명확하게 영향요인을 규명할 필요가 있겠다.

뇌졸중 진단시기에 관한 하현근(2008)의 연구에서 뇌졸중 진단 당시 연령이 낮을수록 HRQoL이 높았다고 보고하였으며, 박주연(2011)의 연구에서 성인의 자살생각과 HRQoL이 음의 상관관계를 나타낸다고 보고하였다. 본 연구의 다변량 분석결과 뇌졸중 진단시기와 자살생각여부가 영향요인으로 확인되지 않았지만, 단변량 분석에서는 그룹간 유의한 차이가 있는 것으로 확인되었다. 이는 선행연구에서는 HRQoL을 연속변수로서 영향요인을 파악한 반면, 본 연구에서는 HRQoL을 이분화하여 영향요인을 파악한 점에서 차이를 나타내었다고 판단된다.

일부 지역사회 노인의 HRQoL 관련요인에 관한 연구에서 치아상태와 HRQoL이 서로 유의한 관련이 있다고 보고된 바 있으며, 본인 치아나 틀니를 착용하고 있는 경우 치아가 없는 경우에 비해 HRQoL이 높았는데, 이는 치아상태가 외모와 관련될 뿐만 아니라 식사 및 대화와 관련하여 사회적 관계와 생활의 수준에 영향을 미치기 때문이다(J. Lim, et al., 2007). 또한 노인은 치아상실이 많아 저작능력이 저하됨으로써 식사의 양과 질이 떨어져서 건강유지에 지장을 초래하게 된다고 하였다(이현옥 & 김진, 2008). 많

은 선행연구에서 구강건강과 관련된 삶의 질이 높을수록 HRQoL이 높아 서로 관련성이 있는 것으로 나타났다(류경자 & 이태용, 2008; 이미라, 2010). 비록 본 연구에서는 하악보철물 상태가 노인 만성질환자의 HRQoL에 영향을 미치는 요인으로 확인되지는 않았으나, 단변량 분석에서는 그룹 간 유의한 차이를 보였으며, 구강건강상태가 HRQoL에 영향을 미친다는 선행연구 결과를 고려한다면 향후 이를 규명하기 위한 더욱 체계적인 연구가 필요할 것으로 생각된다.

연구결과와 관련된 본 연구의 제한점은 다음과 같다.

첫째, 모델 개발을 위한 변수 선정 시에 HRQoL과 직접적인 연관성이 있을 것이라 생각되는 변수들을 제외하는 방법을 고려할 필요가 있다. 예를 들면 주관적 건강상태, 활동제한 여부 및 사유 등의 변수가 모델에 미치는 영향 및 다른 요인들과의 관계성을 파악하여 실제 영향변수인지 혼란변수인지 정확한 판단이 필요하다.

둘째, 본 연구에서는 개별 변수들이 노인 만성질환자의 HRQoL에 미치는 영향 여부만을 제한적으로 살펴보았다. 추가적인 연구로서 변수들이 서로에게 영향을 어떻게 미치는지 변수들간의 관련성뿐만 아니라 직접적인 영향요인인지 다른 요인을 매개로 하는 간접적인 영향요인인지를 확인할 필요가 있겠다.

셋째, ML 적용결과 예측에 기여하는 중요변수들이 정말 유의한 영향요인인지 그 여부를 판단하기 위해서 현재로써는 전통적인 통계적 검증방법에 의존할 수 밖에 없는 한계점을 지닌다. 이에 ML에서 도출된 결과를 합리적으로 검증하고 활용할 수 있는 방법에 대한 후속연구가 필요하다.

넷째, 본 연구는 일반 인구집단을 대상으로 하는 일반도구(General instrument)인 EQ-5D 도구로 측정하여 노인 만성질환자의 HRQoL 그룹을 구분하였기 때문에 포괄적인 HRQoL과 관련된 영향요인들만으로 예측이 가능한 것으로 판단된다. 현재의 EQ-5D 도구는 전반적인 건강상태에 초점이 맞추어진 것이므로 노령화 사회를 위한 보건복지 정책을 위해서는 만성질환을 가진 노인을 대상으로 한 HRQoL 측정도구 개발 및 적용이 필

요하다. 여기서의 측정도구란 보다 복잡한 만성질환의 적응기전을 반영한 질병 특이적 도구(Disease-specific instrument)를 의미하며, 개별 만성질환과 복합 만성질환에 관련된 영향을 모두 포함할 필요가 있다.

ML 기법을 통해 도출된 기여변수들이 정말 유의한 영향변수들인지 판단하기 위해서는 보건의료 전문가들의 타당성 검증을 포함한 보다 구체적인 분석이 요구된다. 기술적인 분석을 통해 기존에 간과하였던 잠재적인 영향요인을 도출하고, 이 결과를 전문가를 통해 검증, 보완하는 과정을 지속적으로 반복할 필요가 있다.

2. 연구방법에 대한 논의

본 연구의 예측모델에서는 SLR 방법이 가장 우수하였으며, ML 기법을 적용한 예측모델은 다소 저하된 성능을 나타내었지만, KNHANES와 같은 혼합 형태(Mixed type)의 자료를 ML 기법을 적용함에 있어서 변수 선택 및 자료 정규화(Data Normalization) 과정을 통해 예측모델의 성능을 향상시킬 수 있음을 알 수 있었다. 다양한 형태의 자료들을 대상으로 한 ML 기법의 적용을 위해서는 Data 속성(형태, 분포 등)을 고려하여 보다 적절한 ML 기법을 적용해야 하며, 변수 선택과 자료 정규화(Data Normalization) 방법을 통해 ML 기법의 성능과 효율성을 향상시킬 수 있다.

본 연구에서는 SLR과 계층적 군집분석을 통한 변수 선정 방법과 Z-transformation, 1-of-K encoding, Discretization이라는 자료 정규화 방법을 사용하였으며, 이를 위해서는 자료에 대한 도메인 전문가 지식 및 통계적 지식이 함께 요구된다. 또한 향후 지속적인 연구와 검증 과정이 요구되며, 유사 형태의 자료를 활용하는 경우에 변수를 선정하고, 자료를 정규화하는 방법에 대한 구체적인 가이드라인을 제시할 수 있을 것이다.

연구방법에 관한 본 연구의 제한점은 다음과 같다.

첫째, EQ-5D index를 두 그룹으로 나누기 위한 기준 즉, Cut-off value

에 대한 문제를 본 연구에서는 기존 문헌의 제시 기준을 근거로 하여 EQ-5D 설문 5문항 모두에 ‘다소 지장이 있다’고 대답한 합산 점수에 해당하는 EQ-5D index 0.678로 정하였으나, 통계적 관점에서 보다 적절한 방안을 제시할 필요가 있다. 연구에서 사용된 자료들의 분포가 정규분포가 아닌 베타분포(skewed distribution) 형태라서 결정이 더 어렵기 때문에 어떤 기준을 선택해야 가장 좋은 예측모델을 만들 수 있을지 통계적 관점에서 전문가 조언을 통해 보다 근본적으로 문제를 해결해야 할 것이다.

둘째, 본 연구에서는 분류 그룹의 sample 수가 매우 큰 차이를 나타내었다. 즉 HRQoL이 양호한 그룹이 그렇지 않은 그룹의 약 10배에 가까운 빈도수를 나타내었는데 이를 보정하기 위해서 DT와 RF 모델에서는 COST function을 사용하였고, SVM의 경우는 feature weight를 10으로 지정하여 각 모델을 개발하였다. 보정과정 자체가 모델에 구체적으로 어떤 영향을 미치는지 알 수 없기 때문에 향후 연구에서는 실제 sample수를 충분히 확보하여 이러한 보정과정 없이 예측모델을 개발할 필요가 있다.

본 연구와 같은 예측모델을 구성함에 있어서 ML 기법을 적용한 최초의 연구로서 전통적으로 시행되던 사회과학적 통계방법 외에도 전산학 분야의 ML 기법을 적용하여 기존 연구와 상호보완적인 측면에서 그 결과를 활용할 수 있는 가능성을 제시하였다는 점에서 본 연구의 의의가 있다. 자료의 설계 또한 기존의 통계적 방법에 맞게 고안되었기 때문에 새로운 방법에는 맞지 않는 형태일 수도 있다. 이런 이유로 ML 기법들은 분석에 대한 성능이 기존 방법을 능가하기 어려울 수도 있지만(Kim, et al., 2011) 빅데이터에 적합한 새로운 기법을 적용함으로써 보다 성능을 높이기 위한 방법들을 이제 찾기 시작했다는 점에서 연구의 가능성을 열어두고 싶다.

기존의 전통적인 접근방식인 통계적 방법은 전문가의 해석과 조정이 필요한 것으로서, 오랜 기간 사용되어 오면서 최적의 성능을 낼 수 있도록 지속적으로 발전되어왔다. 또한 전통적인 통계방법들은 대용량 자료를 다루기 어려운 한계 때문에 전문가의 판단에 의해 정제된 자료를 기반으로 분석을 시도하게 되어 기존의 지식과 유사한 결과를 보여줄 수 밖에 없을 것이

다. 그러나 대용량 자료에 유용한 ML 기법의 경우는 이러한 정제나 해석 없이 기계적으로 분석을 시도하기 때문에 비록 추후 검증이 필요하긴 하지만 기존의 분석 결과와는 다른 새로운 영향요인들을 찾아줄 수 있다. 본 연구는 수많은 분석방법들 중에서 ML 기법에 속하는 몇 가지 방법들을 적용하여, 그 중 가장 좋은 성능을 낼 수 있는 방법을 찾아보았으며, 본 연구를 출발점으로 또 다른 방법을 찾아 분석을 시도하는 후속 연구들을 기대해본다.

3. 연구의 활용 방안

3.1 학문적 의의

우선 보건학적 측면에서 살펴보면, 본 연구결과를 바탕으로 노인 만성질환자의 HRQoL의 영향요인 및 예측모델을 위한 후속 연구들이 진행될 수 있으며, 기존 연구에서 도출된 결과와 본 연구에서 도출된 영향요인을 고려하여 노인 만성질환자의 HRQoL 향상을 위한 맞춤형 건강관리 전략 및 중재프로그램 개발을 위한 적용방안을 다음과 같이 제안할 수 있다.

만성질환을 가진 노인들을 대상으로 효율적인 건강관리를 할 수 있도록 경제적인 지원이 우선적으로 필요하며, 이는 만성질환 관리료의 형태로 일정금액을 개인별 차등화하여 지원하거나 노인 만성질환 관리를 위한 전문 시설과 전문인력을 확충하여 지역사회를 통한 사회복지 서비스를 제공하는 방안을 마련할 수 있을 것이다.

노인 만성질환자에게 적합한 질병 특이적 HRQoL 측정도구를 개발하여 적용함으로써 정확하고 구체적인 평가지표들을 확인할 수 있을 것이다. 이를 통해 만성질환 유형별로 특성에 맞추어 HRQoL 증진방안을 마련하고, 만성질환에 의한 우울 및 불편감 완화를 위한 친목모임과 자원봉사활동 등 사회참여 활성화 서비스를 개발하여 제공할 필요가 있겠다. 또한 노인 스스로 자신의 만성질환을 수용하고 잘 적응하도록 돕고, 만성질환을 가지고도 행복하고 건강한 삶을 영위할 수 있도록 실제적인 성공사례와 적응방법들

에 대한 구체적인 정보를 제공하며, 서로 공유할 수 있는 사회 연계망을 구축할 필요가 있다.

의료정보학적 측면에서는 ML 기법을 활용하는 경우 보건의료 환경에서 얻어지는 유사자료들을 대상으로 한 변수 선정 및 자료 정규화 방법에 대한 연구를 통해 구체적인 가이드라인 제시가 가능하다. 본 연구에서는 KNHANES 자료에 대하여 SLR과 계층적 군집분석을 통한 변수 선정 방법과 Z-변형(Z-transformation), 이진화(1-of-K encoding scheme coding), 이산화(Discretization), 신규 변수 생성이라는 자료 정규화 방법을 구체적으로 제시하였다. 본 연구 방법을 활용한 후속 연구들을 통해서 재검증하고, 수정 및 보완 함으로써 발전적인 방향으로 가이드라인이 더욱 견고해질 것이다. 이외에도 많은 다양한 자료들을 다루는 시도들이 계속되고, 그 결과를 함께 공유함으로써 수많은 가이드라인이 생명력을 가지고 진화할 수 있기를 기대해본다.

3.2 실무에의 적용

본 연구의 방법 및 결과를 활용하여 다음과 같은 시스템을 제안한다. 전자 의무기록(Electronic Medical Record: 이하 EMR) 또는 개인건강기록(Personal Health Record: 이하 PHR)과 같은 Data repository에서 Data extraction & Feature filtering system을 통해 다양한 요인들을 포함하여 HRQoL 분석과 예측을 위한 Feature set을 얻어낼 수 있다. 이후 Machine learning analysis program을 적용하여 얻은 분석 결과를 Health manager가 실시간으로 또는 원하는 경우 수시로 전달받게 된다. Health manager는 분석결과에 적합한 Health promotion prescription을 내리고 HRQoL 위험군을 대상으로 건강증진행위를 실천할 수 있도록 중재하게 된다. 이를 통해 HRQoL 위험군을 대상으로 한 중재와 적용 결과들이 EMR 또는 PHR로 다시 입력되어 실시간 반복적으로 HRQoL을 분석하고 예측하는 과정을 되풀이하게 된다.

예를 들면, 제안 시스템이 Health manager에게 대상자가 HRQoL 위험군에 해당되며, 최대 골밀도 연령군의 골밀도 평균에 대한 표준편차를 의미하는 T-score -1.8, 즉 골감소증에 해당된다(질병관리본부, 2011)는 분석결과를 제공하였다면 Health manager는 이에 대한 HRQoL 중재방안으로서 골다공증 예방을 위한 다양한 건강증진 처방을 제공하게 된다. 골다공증이 뼈의 양이 감소하고 질적인 변화로 인해 뼈의 강도가 약해져 골절의 위험성이 높아지는 질환으로서 이를 예방하기 위해서는 비타민 D와 칼슘, 인의 흡수를 돕는 적절한 식습관과 규칙적인 실외운동 등이 도움을 주며, 골절이 쉽게 되기 때문에 낙상방지를 위한 주의가 더욱 필요하다. 따라서 HRQoL 위험군을 대상으로 골다공증 예방을 위한 집중적인 교육 및 중재를 제공한다면, 결과적으로 HRQoL이 향상되어 위험군에서 벗어날 수 있게 된다. 본 연구에서 제안하는 전체 시스템 구조는 다음과 같다(그림 26).

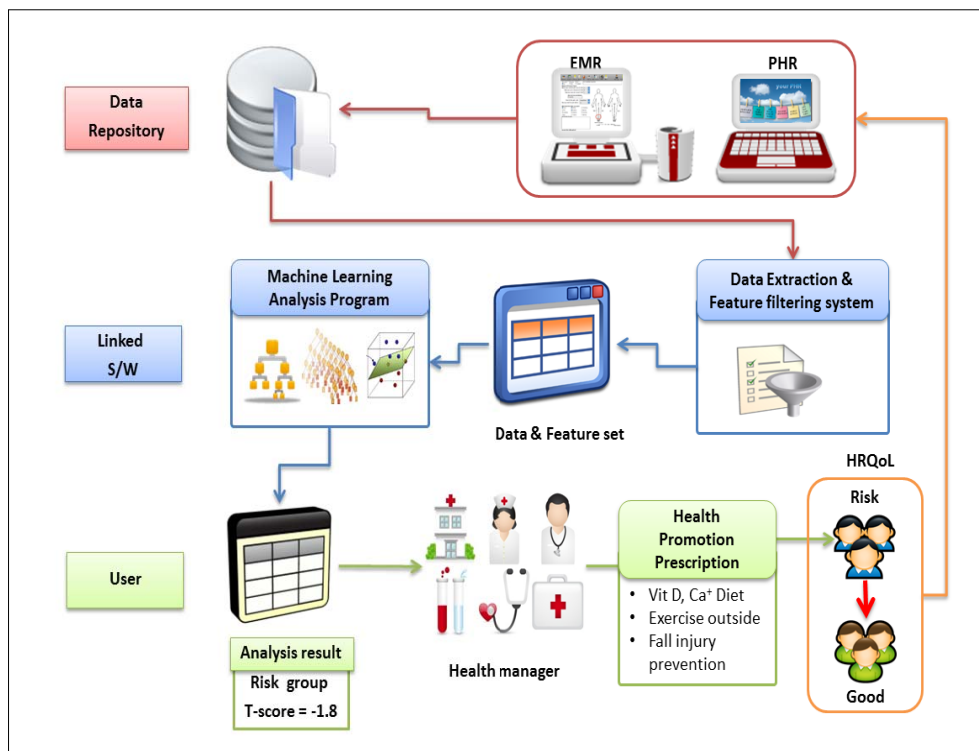


Figure 26. Proposed system architecture

의료기관의 EMR 환경에서 ML의 source가 되는 features를 온톨로지로 구성하여 임상현장에서 실시간으로 변수들에 관한 정보 및 사례들을 모아주는 시스템을 개발한다면 의료전문가들에게 알려진 지식들 외에 새로운 지식들을 제공할 수 있을 것이다. 기존의 임상 의사결정지원시스템(Clinical Decision Support System: 이하 CDSS)이 이미 정해진 규칙에 의한 정보를 제공한다면 Point of care로서 ML 프로그램은 새로운 지식을 제공함으로써 상호보완적 역할을 할 것이다.

또한 의료전문가가 이미 알고 있는 지식 범주에서 판단할 수 있는 내용을 넘어서는 보다 복잡한 질병치료의 경우, 알려지지 않았으나 잠재적으로 중요한 변수들을 예측함으로써 환자의 질병치료에 즉시적으로 도움을 줄 수 있으며, 이러한 지식들이 쌓이면 Critical pathway를 예측하여 제시할 수 있는 연관 시스템도 개발될 수 있을 것이다.

VII. 결론 및 제언

1. 결론

본 연구는 KNHANES 자료를 활용하여 노인 만성질환자를 대상으로 한 HRQoL 영향요인을 분석하고, HRQoL 위험군 예측모델을 개발하기 위하여 전통적인 SLR 방법과 새로운 ML 기법을 적용하였다.

본 연구에서 노인 만성질환자의 HRQoL에 영향을 미치는 요인들로 ‘소득’, ‘만성질환 종류’, ‘우울’, ‘불편감’, ‘주관적인 건강상태’가 해당되었다. 즉, 소득에 따라, 만성질환 종류에 따라, 2주 이상 연속 우울감 여부에 따라, 최근 2주간 불편감 일수에 따라, 주관적인 건강상태에 따라 노인 만성질환자의 HRQoL이 달라질 수 있음을 알 수 있었다. ML 적용결과 노인 만성질환자의 HRQoL 예측을 위한 중요변수로서 ‘골관절염 소요기간’, ‘골다공증 유병여부’, ‘뇌졸중 진단시기’, ‘하악보철물 상태’, ‘1년간 자살생각여부’ 등이 추가로 도출되었다. 이러한 요인들은 HRQoL 예측에 영향을 미치는 기여변수들이나 통계적으로 검증되지 않았기 때문에 가설 모형에서는 제외하였다.

구축된 예측모델의 타당도 검증은 10 fold cross-validation 방법을 적용하였으며, 성능 평가를 위해서는 Accuracy, F-score 등의 값을 모형별로 산출하여 비교한 결과 SLR이 Accuracy=0.93, F-score=0.49로 가장 우수한 성능을 나타내었으며, 그 다음으로는 SVM (0.90, 0.51), RF (0.87, 0.33), DT (0.82, 0.23) 순이었다.

본 연구의 예측모델에서는 ML의 3가지 방법들이 SLR 방법보다 다소 저하된 성능을 나타내었지만, 전통적인 방법과 새로운 방법의 상호보완적인 적용을 통해 새로운 영향요인을 확인할 수 있었으며, KNHANES와 같은 혼합 형태(Mixed type)의 자료에 ML 기법을 적용함에 있어서 변수 선정과 자료 정규화(Data Normalization) 과정을 통해 예측모델의 성능과 효율을

향상시킬 수 있음을 알 수 있었다. 향후 이에 대한 지속적인 연구와 검증 과정이 요구되며, 유사 형태의 자료를 활용하는 경우에 자료를 정규화하는 방법에 대한 구체적인 가이드라인을 제시할 수 있을 것이다. 본 연구와 같이 의료분야에서의 자료 재창출(Data Repositioning)을 통해 가치있는 새로운 지식을 도출할 수 있도록 이러한 접근 방법이 더욱 활발하게 적용되기를 기대한다.

2. 제언

본 연구는 노인 만성질환자들의 HRQoL을 예측하고, 평가함으로써 이를 향상시키기 위한 구체적이고 합리적인 계획을 마련할 수 있도록 돕고자 시도되었다. 이상의 결론을 근거로 다음과 같이 제언한다.

첫째, 노인의 만성질환 적응 및 대응기전과 관련된 HRQoL 영향요인에 대한 후속 연구가 필요하며, 만성질환을 가진 노인을 대상으로 한 질병 특이적 HRQoL 도구 개발 및 적용에 관한 연구가 병행되어야 할 것이다.

둘째, HRQoL 분류시 2집단보다는 3집단 또는 4집단으로 분류하는 것이 결과 해석력 및 적용 가능성이 높다고 할 수 있다. 또한 EQ-5D 외에 LQ_VAS(오늘 당신의 건강상태 0~100점) 등의 관련 변수를 추가해서 다각도의 분석이 필요하다. 현재의 단순한 예측모델이 아닌 다차원 예측모델로 확장하는 방법을 고려해야 한다. 즉, HRQoL을 보다 다양한 카테고리로 나누어서 분석할 필요가 있다.

셋째, ML 기법에 적합한 자료구조를 찾아 모델을 개선한다면 더 좋은 결과를 이끌어 낼 수 있을 것이다. 본 연구에서는 기계학습을 위해 1-of-K encoding 등을 포함하여 자료 정규화를 시행하였으나, 최적화된 Data transforming을 위해 이산화 등 더 폭넓은 다양한 방법을 시도해 볼 필요가 있다. 이에 기계학습에 적합한 자료 형태 및 구조를 밝히는 후속 연구를 제안한다.

넷째, 본 연구는 주로 대상자의 설문과 검진자료로 구성된 변수들을 사용하였으나, EMR 또는 PHR 등의 실제적인 Health record를 통해 보다 다양한 변수들을 추출하여 예측모델을 구성할 필요가 있다. 이를 위해서는 관련 Document 또는 Database로부터 어떻게 자료를 추출할 것인가에 대한 연구가 함께 이루어져야 할 것이다.

최근 신약 재창출(Drug repositioning) 연구가 활발한데, 신약 재창출이란 충분한 안전성과 선행 데이터를 가지고 있는 약물을 새로운 적응증을 타겟으로 다시 개발하는 것을 의미하는 것으로 새로운 신약개발의 방법으로 등장하고 있다. 이는 재발견 또는 재활용이라는 관점에서 다른 분야에도 적용될 수 있으며, 바로 의료분야에서의 자료 재창출(Data repositioning)을 통해 가치있는 새로운 지식을 도출해낼 수 있음을 의미한다. 기존에는 알지 못했던 새로운 지식을 발견함으로써 실제적인 임상현장에서의 재활용을 가능하게 한다.

결론적으로 보다 유용한 자료를 활용하는 것이 성공적인 예측모델을 개발할 수 있는 중요한 열쇠라고 생각된다. 이러한 예측모델을 통해 자료의 재생성 및 관리가 이루어지고, 변화된 자료를 통해 더 좋은 예측모델을 개발할 수 있을 것이다. 즉, Data repositioning이라는 trigger event를 통해 효율적인 예측모델의 개발을 기대할 수 있을 것이다. 또한 복잡하고 불명확한 전문가의 지식을 보다 구체화하고, 향상시키기 위해서는 이러한 자료 재창출 과정들이 반복됨으로써 결국은 최상의 질을 가진 데이터로 점차 변화되고, 그 결과 전문가 지식의 확장과 개선을 통해 HRQoL이 향상되는 선순환적인 과정이 필요하다(그림 27).

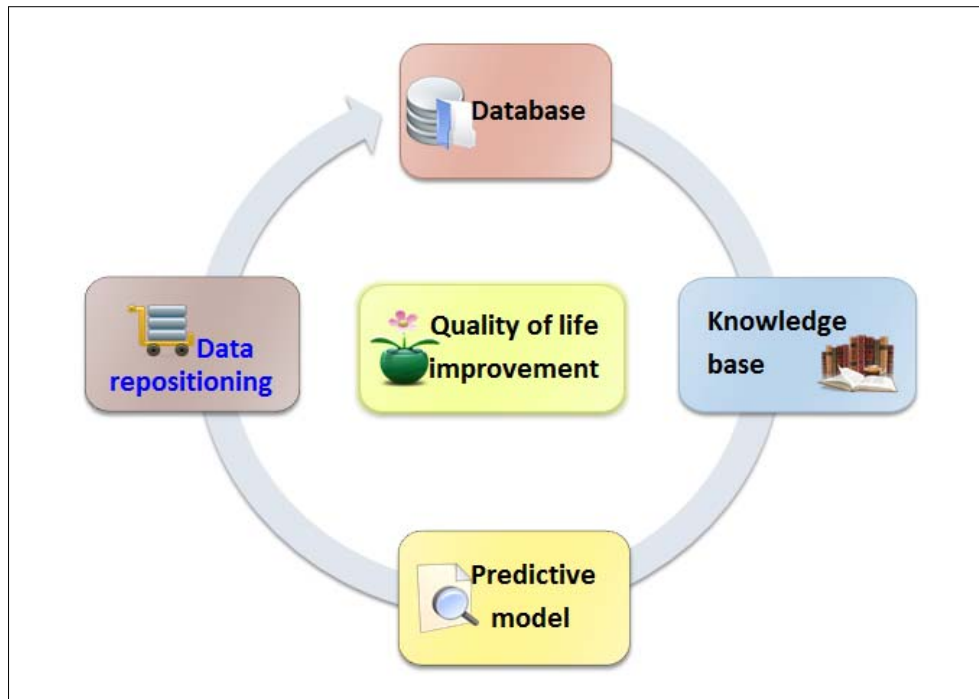


Figure 27. Virtuous cycle of data repositioning

본 연구가 노인 만성질환자의 HRQoL 분석에 대하여 기존에 적용하지 않았던 새로운 방법을 통해 HRQoL을 바라보는 새로운 관점을 제시하는 가치 있는 첫걸음이 되기를 기대해본다.

References

- Ahn, S. H., Park, B. W., Noh, D. Y., Nam, S. J., Lee, E. S., Lee, M. K., et al. (2007). Health-related quality of life in disease-free survivors of breast cancer with the general population. *Annals of Oncology*, 18(1), 173-182.
- Alguren, B., Fridlund, B., Cieza, A., Sunnerhagen, K. S., & Christensson, L. (2012). Factors Associated With Health-Related Quality of Life After Stroke: A 1-Year Prospective Cohort Study. *Neurorehabilitation and neural repair*, 26(3), 266-274.
- Allison, D. B., Cui, X., Page, G. P., & Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. Review]. *Nat Rev Genet*, 7(1), 55-65.
- Anooj, P. K. (2011). Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules. *Central European Journal of Computer Science*, 1(4), 482-498.
- Austin, P. C., Lee, D. S., Steyerberg, E. W., & Tu, J. V. (2012). Regression trees for predicting mortality in patients with cardiovascular disease: What improvement is achieved by using ensemble-based methods? *Biometrical Journal*, 54(5), 657-673.
- Bellos, C., Papadopoulos, A., Rosso, R., & Fotiadis, D. (2012). *Categorization of patients' health status in COPD disease*

using a wearable platform and random forests methodology.

Paper presented at the Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on.

Benedict, R. H. B., Wahlig, E., Bakshi, R., Fishman, I., Munschauer, F., Zivadinov, R., et al. (2005). Predicting quality of life in multiple sclerosis: accounting for physical disability, fatigue, cognition, mood disorder, personality, and behavior change. *Journal of the neurological sciences*, 231(1), 29-34.

Berton, H. K., Cassel, J. C., & Gore, S. (1977). Social Support and Health. *Medical Care*, 15(5), 47-58.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4): springer New York.

Bowling, A., Gabriel, Z., Dykes, J., Dowding, L. M., Evans, O., Fleissig, A., et al. (2003). Let's ask them: a national survey of definitions of quality of life and its enhancement among people aged 65 and over. *The International Journal of Aging and Human Development*, 56(4), 269-306.

Brown, E. A. (2012). What can we do to improve quality of life for the elderly chronic kidney disease patient? *Aging Health*, 8(5), 519-524.

Burckhardt, C. S. (1985). The impact of arthritis on quality of life. *Nursing Research*.

Cao, Y., Xu, J., Liu, T. Y., Li, H., Huang, Y., & Hon, H. W. (2006). *Adapting ranking SVM to document retrieval*. Paper presented at the Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.

Chan, K., Lee, T. W., Sample, P. A., Goldbaum, M. H., Weinreb, R. N.,

- & Sejnowski, T. J. (2002). Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *Biomedical Engineering, IEEE Transactions on*, 49(9), 963-974.
- Cheadle, C., Vawter, M. P., Freed, W. J., & Becker, K. G. (2003). Analysis of microarray data using Z score transformation. *The Journal of molecular diagnostics*, 5(2), 73-81.
- Chen, J., Xing, Y., Xi, G., Chen, J., Yi, J., Zhao, D., et al. (2007). A Comparison of Four Data Mining Models: Bayes, Neural Network, SVM and Decision Trees in Identifying Syndromes in Coronary Heart Disease. *Advances in Neural Networks- ISNN 2007*, 1274-1279.
- Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics*, 17(1), 13-35.
- Cusano, C., Ciocca, G., & Schettini, R. (2003). *Image annotation using SVM*. Paper presented at the Electronic Imaging 2004.
- Daggett, V., Bakas, T., & Habermann, B. (2009). A review of health-related quality of life in adult traumatic brain injury survivors in the context of combat veterans. *Journal of Neuroscience Nursing*, 41(2), 59-71.
- Duda, R. (2002). *Pattern Classification*: Wiley Interscience.
- Ferrans, C. E., Zerwic, J. J., Wilbur, J. E., & Larson, J. L. (2005). Conceptual Model of Health-Related Quality of Life. *Journal of Nursing Scholarship*, 37(4), 336-342.
- Fung, G., & Stoeckel, J. (2007). SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. *Knowledge and Information Systems*, 11(2), 243-258.

- Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., & Rueckert, D. (2012). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*.
- Groessl, E. J., Kaplan, R. M., Barrett-Connor, E., & Ganiats, T. G. (2004). Body mass index and quality of well-being in a community of older adults. *American journal of preventive medicine*, 26(2), 126.
- Group, T. E. (1990). EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199-208.
- Höfer, S., Benzer, W., Alber, H., Ruttman, E., Kopp, M., Schussler, G., et al. (2005). Determinants of health-related quality of life in coronary artery disease patients: a prospective study generating a structural equation model. *Psychosomatics*, 46(3), 212-223.
- Han, K. S. (2003). Perceived stress, mood state, and symptoms of stress of the patient with chronic illness. *Journal of Korean academy of nursing*, 33(1), 87-94.
- Heckman, T. G. (2003). The chronic illness quality of life (CIQOL) model: Explaining life satisfaction in people living with HIV disease. *Health Psychology*, 22(2), 140.
- Holmes, S. (1985). Research. Pursuit of happiness. *Nursing mirror*, 161(13), 43.
- IBM. (2012). SPSS version 20.0. Retrieved 2012.12.26, from IBM Corp.: <http://spss.co.kr/>
- J Taylor, S. L., Kathryn M McPherson, Harry K McNaughton, W. (2001). EuroQol EQ-5D may not adequately describe the health of people with disabilities. *Disability & Rehabilitation*, 23(7), 281-285.

- Jiang, Y., & Hesser, J. E. (2006). Associations between health-related quality of life and demographics and health risks. Results from Rhode Island's 2002 behavioral risk factor survey. *Health and quality of life Outcomes*, 4(1), 14.
- Joung, K. H. (2011). Health-related quality of Life among home-dwelling people with arthritis in Korea: Comparative study of osteoarthritis and rheumatoid arthritis. *Journal of the Korean Data & Information Science Society*, 22(3), 555-563.
- Kang, E., Kim, N., Kim, D., Kim, H., Beon, Y., & Seo, M. (2007). *Korean health-related quality of life and health-adjusted life expectancy*. Seoul: Korea Centers for Disease Control and Korea Institute for Health and Social Affairs.
- KCDC. (2010). *The Fifth Korea National Health and Nutrition Examination Survey*.
- KCDC. (2012). Introduction of KNHANES. Retrieved 2012.12.26, from Centers for Disease Control and Prevention: <http://knhanes.cdc.go.kr/knhanes/index.do>
- Kempen, G., Ormel, J., Brilman, E. I., & Relyveld, J. (1997). Adaptive responses among Dutch elderly: the impact of eight chronic medical conditions on health-related quality of life. *American Journal of Public Health*, 87(1), 38-44.
- Kim, S., Kim, W., & Park, R. W. (2011). A Comparison of Intensive Care Unit Mortality Prediction Models through the Use of Data Mining Techniques. *Healthcare informatics research*, 17(4), 232-243.
- Laborde, J. M., & Powers, M. J. (1985). Life satisfaction, health control orientation, and illness-related factors in persons with osteoarthritis. *Research in nursing & health*, 8(2), 183-190.

- Larson, R. (1978). Thirty years of research on the subjective well-being of older americans. *J Gerontol*, 33(1), 109-125.
- Lepetit, V., & Fua, P. (2006). Keypoint recognition using randomized trees. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9), 1465-1479.
- Lim, J., Park, J., Kang, M. G., & Ryu, S. Y. (2007). Quality of Life and its Associated Factors among Some Elderly Residents Using a Hall for the Aged in a Community. *J Prev Med Public Health*, 40(5), 337-344.
- Lim, Y. T., Park, Y. W., Kim, C. H., Shin, H. C., & Kim, J. W. (2001). Effect of weight loss on health related quality of life in obese patients. *Journal of the Korean Academy of Family Medicine*, 22(4), 556-564.
- Luengo, J., Saez, J., Lopez, V., & Herrera, F. (2012). A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning.
- Lunetta, K. L., Hayward, L. B., Segal, J., & Van Eerdewegh, P. (2004). Screening large-scale association study data: exploiting interactions using random forests. *BMC Genet*, 5, 32.
- Marchesini, G., Bianchi, G., Amodio, P., Salerno, F., Merli, M., Panella, C., et al. (2001). Factors associated with poor health-related quality of life of patients with cirrhosis. *Gastroenterology*, 120(1), 170-178.
- Marquis, M. S., Davies, A. R., & Ware, J. E., Jr. (1983). Patient satisfaction and change in medical care provider: a longitudinal study. [Research Support, U.S. Gov't, P.H.S.]. *Med Care*, 21(8), 821-829.
- Maruta, T., Vatterott, M. K., & McHardy, M. J. (1989). Pain

- management as an antidepressant: long-term resolution of pain-associated depression. *Pain*, 36(3), 335-337.
- MathWorks. (2012). MATLAB Revised for Version 7.14. Retrieved 2012.12.26, from The MathWorks, Inc.:
<http://www.mathworks.co.kr/>
- Mitchell, T. M. (1997). *Machine Learning*: The McGraw-Hill Companies.
- Nam, H., Kim, K., Kwon, S., Koh, K., & Poul, K. (2007). EQ-5D Korean valuation study using time trade off method. *Seoul: Korea Centers for Disease Control and Prevention*.
- NLM. (2012). Medical Subject Headings. from National Library of Medicine
http://www.nlm.nih.gov/cgi/mesh/2013/MB_cgi?mode=&index=11283&view=expanded
- Ocampo, J. M. (2010). Self-rated health: Importance of use in elderly adults. *Colombia Médica*, 41(3), 275-289.
- Oh, H. S. (1993). Health promoting behaviors and quality of life of Korean women with arthritis. *The Journal of Nurses Academic Society*, 23(4), 617-630.
- Olanow, C. W., Watts, R. L., & Koller, W. C. (2001). An algorithm (decision tree) for the management of Parkinson's disease (2001): Treatment Guidelines. *Neurology*, 56(11), S1-S88.
- Oleson, M. (1990). Subjectively perceived quality of life. *Image J Nurs Sch*, 22(3), 187-190.
- Park, Y. W., Shin, H. C., & Kim, C. H. (2000). Health-related quality of life in people with overweigh and large waist circumference. *Journal of the Korean Academy of Family Medicine*, 21(6), 753-761.
- Peruzza, S., Sergi, G., Vianello, A., Pisent, C., Tiozzo, F., Manzan, A.,

- et al. (2003). Chronic obstructive pulmonary disease (COPD) in elderly subjects: impact on functional status and quality of life. *Respiratory medicine*, 97(6), 612-617.
- Picavet, H. S., & Hoeymans, N. (2004). Health related quality of life in multiple musculoskeletal diseases: SF-36 and EQ-5D in the DMC3 study. [Research Support, Non-U.S. Gov't]. *Ann Rheum Dis*, 63(6), 723-729.
- Polat, K., Güneş, S., & Arslan, A. (2008). A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine. *Expert Systems with Applications*, 34(1), 482-487.
- Porojan, M., Poanta, L., & Dumitrascu, D. L. (2012). Assessing health related quality of life in diabetic patients. *Romanian journal of internal medicine*, 50(1), 27-32.
- Ropka, M. E. (2012). Symptom status and functional status outcomes: humanistic outcomes in obesity disease management. *Obesity research*, 10(S1), 42S-49S.
- Rubin, R. R., & Peyrot, M. (1999). Quality of life and diabetes. *Diabetes/metabolism research and reviews*, 15(3), 205-218.
- Sajid, M., Tonsi, A., & Baig, M. (2008). Health-related quality of life measurement. *International journal of health care quality assurance*, 21(4), 365-373.
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210-229.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L.,

- Aguiar, R. C. T., et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature medicine*, 8(1), 68-74.
- Solomonoff, A., Quillen, C., & Campbell, W. M. (2004). *Channel compensation for SVM speaker recognition*. Paper presented at the Proceedings of Odyssey.
- Son, Y. J., Kim, H. G., Kim, E. H., Choi, S., & Lee, S. K. (2010). Application of Support Vector Machine for Prediction of Medication Adherence in Heart Failure Patients. *Healthcare informatics research*, 16(4), 253-259.
- Sousa, K. H., & Kwok, O. M. (2006). Putting Wilson and Cleary to the test: analysis of a HRQOL conceptual model using structural equation modeling. *Quality of Life Research*, 15(4), 725-737.
- Spilker, B. (1996). *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia: Lippincott-Raven.
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. [Research Support, Non-U.S. Gov't]. *BMC Bioinformatics*, 8, 25.
- Stromgren, A. S., Goldschmidt, D., Groenvold, M., Petersen, M. A., Jensen, P. T., Pedersen, L., et al. (2002). Self-assessment in cancer patients referred to palliative care - A study of feasibility and symptom epidemiology. *Cancer*, 94(2), 512-520.
- Stuifbergen, A. K., & Becker, H. A. (1994). Predictors of health-promoting lifestyles in persons with disabilities. *Res Nurs Health*, 17(1), 3-13.
- Stuifbergen, A. K. (1995). Health-promoting behaviors and quality of life among individuals with multiple sclerosis. *Research and*

Theory for Nursing Practice, 9(1), 31-50.

Stuifbergen, A. K., & Rogers, S. (1997). Health promotion: An essential component of rehabilitation for persons with chronic disabling conditions. *Advances in Nursing Science*, 19(4), 1-20.

Stuifbergen, A. K., Seraphine, A., Harrison, T., & Adachi, E. (2005). An explanatory model of health promotion and quality of life for persons with post-polio syndrome. *Social Science & Medicine: Social Science & Medicine*.

Summer, L., O'Neill, G., & Shirey, L. (1999). *Chronic conditions: A challenge for the 21st century*: National Academy on an Aging Society.

Takahashi, N. (2012). Efficient Learning Algorithms for Support Vector Machines Retrieved 2012.12.26:
<http://www-kairo.csce.kyushu-u.ac.jp/~norikazu/research.en.html>

van der Zanden, B. P., Dijkgraaf, M. G. W., Blanken, P., de Borgie, C. A. J. M., van Ree, J. M., & van den Brink, W. (2006). Validity of the EQ-5D as a generic health outcome instrument in a heroin-dependent population. *Drug and Alcohol Dependence*, 82(2), 111-118.

Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44(2), 330-349.

Ward, J. J., McGuffin, L. J., Buxton, B. F., & Jones, D. T. (2003). Secondary structure prediction with support vector machines. [Comparative StudyEvaluation Studies Research Support, Non-U.S. Gov't Validation Studies]. *Bioinformatics*, 19(13),

1650-1655.

- Ware, J., Kosinski, M., Dewey, J., & Gandek, B. (2001). How to score and interpret single-item health status measures: a manual for users of the SF-8 health survey. Lincoln (RI): QualityMetric: Inc.
- Wettergren, L., Björkholm, M., Axdorph, U., & Langius-Eklöf, A. (2004). Determinants of health-related quality of life in long-term survivors of Hodgkin's lymphoma. *Quality of Life Research*, 13(8), 1369-1379.
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. *JAMA: the journal of the American Medical Association*, 273(1), 59-65.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.
- Xin, M., & Xiao, G. J. S. (2012). Prediction of RNA-binding residues in proteins using random forest. *Journal of Southeast University (Natural Science Edition)*, 1, 011.
- Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak*, 10, 16.
- Yu, Y., & Wang, X. (2002). Machine Learning for Information Extraction : Approaches and Applications. *Journal of Korean Language Information Science Society*, 7(8), 68-79.
- Zhang, B.-T. (2010). Machine Learning: Samsung Advanced Institute of Technology.
- Zhang, L., Lin, F., & Zhang, B. (2001). *Support vector machine learning for image retrieval*. Paper presented at the Image

Processing, 2001. Proceedings. 2001 International Conference on.

강경숙. (2003). 일 농촌지역 노인의 사회적 지지와 삶의 질. *지역사회간호학회지*, 14(3), 375-384.

건강보험심사평가원. (2012). *2011년 진료비통계지표 개요 보고서*.

고승덕, & 조숙행. (1997). 연구논문 : 노인의 삶의 질 향상을 위한 요인 추출. *한국노년학*, 17(2), 17-36.

길선령, 이상일, 윤성철, 안형미, 조민우, 엄홍대, et al. (2008). 우리나라 성인에서 일부 질환과 연관된 건강관련 삶의 질 감소. *예방의학회지*, 41(6), 434-441.

김광준, & 이형근. (2011). 압타머 단백질 바이오칩을 이용한 간암 진단 생체정보 예측 시스템 개발. *한국전자통신학회논문지*, 6(6).

김동희, 함기백, & 김진. (2006). SVM을 이용한 만성간염 환자 예측진단을 위한 SNP정보분석. *한국정보과학회 학술발표논문집*, 33(1), 19-21.

김명자, & 송경애. (1991). 지지간호가 만성질환자의 삶의질과 자존감에 미치는 효과. *대한간호학회지*, 21(3), 323-338.

김미령. (2006). 전기, 후기 여성노인의 삶의 질 및 영향요인 비교연구. *한국사회복지학*, 58(2), 197-222.

김상윤. (2011). *서포트 벡터 머신과 인공 신경망을 이용한 전립선 조직검사를 통한 전립선암의 병기 예측 모델 설계 및 각 모델 간 정확도 비교*. 강원대학교, 춘천.

김수형. (2007). 문자인식 분야의 기계학습 기법. *정보과학회지*, 25(3), 12-20.

김정연, 이석구, & 이성국. (2010). 노인에서 건강행태, 건강수준, 일상생활수행능력, 건강관련 삶의 질과의 관계. *한국노년학*, 30(2), 471-484.

김태현, 김동배, 김애순, 김미혜, & 이영진. (1999). 노년기 삶의 질 향상에 관한 연구(2). *한국노년학*, 19(1), 61-81.

- 김혜민. (2012). *무릎 골관절염 환자의 건강관련 삶의 질 및 관련 요인*. 아주대학교, 수원.
- 김홍열, & 윤숙영. (2003). 만성질환을 가진 노인의 정서적 안정에 미치는 원예치료의 효과. *화훼연구*, 11(1), 67-72.
- 남상권, & 심옥수. (2011). 만성질환노인의 삶의 질 영향요인 연구: 만성질환유형을 중심으로. *노인복지연구*, 53(단일호), 196-216.
- 남해성, 손명호, 권순석, 김순영, 박형철, 이청우, et al. (2000). 광주광역시 노인들의 만성질환과 건강관련 삶의 질. *노인병학회지* 4(3), 172-190.
- 류경자, & 이태용. (2008). *일부 노인의 구강건강 수준과 구강건강관련 삶의 질의 관련성*. 충남대학교 대학원, 대전.
- 문은영. (2004). *만성질환 노인의 외래 의료서비스 이용 행태에 관한 연구*. 서강대학교 공공정책대학원, 서울.
- 박강희, & 신현정. (2010). 글로벌 경제-금융지표를 이용한 개별 주가예측. *춘계학술대회논문집*, 2010(단일호), 1-30.
- 박상규. (2006). 노인의 정신건강과 삶의 질간의 관계. *한국심리학회지 건강*, 11(4), 785-796.
- 박용섭. (2010). *빈곤층의 삶의 질에 지역사회 복지자원이 미치는 영향*. 연세대학교 대학원, 서울.
- 박은자. (2010). 노인의 만성질환치료제 비용부담. *보건복지포럼*, 170, 51-59.
- 박주연. (2011). *한국 성인의 삶의 질과 자살생각의 관련성*. 중앙대학교, 서울.
- 박혜순, 오상우, 강재현, 박용우, 최중명, 김용성, et al. (2003). 한국인에서 대사증후군의 유병률 및 관련 요인-1998 년 국민건강영양조사를 바탕으로. *대한비만학회지*, 12(1), 1-14.
- 박혜영, & 이관용. (2011). *패턴인식과 기계학습*: 이한출판사.
- 배상열, 고대식, 노지숙, 이병훈, 박형수, & 박종. (2010). 한국 노인의 신체

- 활동과 건강관련 삶의 질의 관계. *한국콘텐츠학회논문지*, 10(10), 255-266.
- 보건산업진흥원. (2007). 노인집단 중심 만성질환관리 지원사업.
- 서동혁, 신동문, 손호선, 김원재, 김원태, & 류근호. (2012). 분류 기법을 이용한 방광암 재발 예측. *정보과학회논문지: 데이터베이스*, 39(3).
- 서성은. (2010). *노인의 건강행위실천과 비만도와의 관련성*. 연세대학교 보건대학원, 서울.
- 성상석, 최찬범, 성윤경, 박용욱, 이혜순, 엄완식, et al. (2004). 한국인에서 EQ-5D 를 이용한 건강 관련 삶의 질 측정. *대한류마티스학회지*, 11(3), 254-262.
- 성은주. (2005). 한국인의 만성질환. *한국유가공기술과학회지*, 23(2), 99-106.
- 손민선, & 문병순. (2012). 빅 데이터 시대의 한국, 갈라파고스가 되지 않으려면. *LG경제연구원*.
- 손신영. (2006). *농촌 노인의 삶의 질 모형구축*. 서울대학교 대학원, 서울.
- 손연정. (2007). 기혼 직장여성의 건강관련 삶의 질에 영향을 미치는 요인. *성인간호학회지*, 19(3), 422-435.
- 송미순. (2004). 건강증진개념을 적용한 만성질환 노인 자기관리 교육 모형 개발. *노인간호학회지*, 6(2), 228-242.
- 송미혜, 이진, 조성필, & 이경중. (2005). SVM 분류기를 통한 심실세동 검출. *전자공학회논문지* 42(5), 27-34.
- 신경림, & 김정선. (2004). 도시지역 여성노인의 건강관심도, 주관적 건강평가, 건강상태 및 건강증진행위와의 관계연구. *대한간호학회지*, 34(5), 869-880.
- 신동규. (2012). Decision Tree 분류기를 사용한 심전도 데이터 정확도 향상에 관한 연구. *한국정보과학회 2012 한국컴퓨터종합학술대회 논문집*, 39(1).
- 신상진. (2010). *한국성인들의 건강관련 삶의 질 특성분석*. 서울대학교 보

건대학원, 서울.

- 신영희, & 이혜정. (2009). 일 도시 노인의 성공적인 노화 관련 요인. *한국노년학*, 29(4), 1327-1340.
- 신지원, 박용경, 서순림, & 김지은. (2011). 한국노인 당뇨병환자의 삶의 질 영향요인. *한국노년학*, 31(3), 479-487.
- 안소현. (2008). *우리나라 노인들의 만성질환과 관련된 식이 위험인자에 관한 연구*. 가톨릭대학교 대학원, 부천.
- 안홍모. (2010). 제 4기 국민건강영양조사에서 한국 성인 남 녀의 생애 주기에 따른 건강관련 삶의 질 관련 요인. 연세대학교 보건대학원, 서울.
- 유재학, 임영희, 이한성, 박대희, & 장진경. (2009). SVM을 이용한 신노년층의 생활 만족도 예측 모형. *한국정보과학회 학술발표논문집*, 36(1), 352-356.
- 이금분. (2010). *다중 클래스 SVM을 이용한 부정맥 신호 분류*. 조선대학교, 광주.
- 이동호, & 빈성오. (2011). 노인들의 질환별 유병여부와 건강관련 삶의 질 구조모형 분석. *한국콘텐츠학회논문지*, 11(1), 216-224.
- 이미라. (2010). *자가 인지한 구강건강상태가 삶의 질에 미치는 영향*. 가천의과학대학교 대학원, 인천.
- 이미숙. (2010). 도시거주 노인의 만성질환과 우울증세의 상관성 및 사회관계망의 효과에 대한 연구. *보건과 사회과학*, 27, 5-30.
- 이병희, & 고주연. (2010). 정신지체아동 및 뇌성마비아동의 건강관련 삶의 질에 미치는 영향 요인 탐색. *특수교육재활과학연구*, 49(2), 105-126.
- 이용철, 임복희, & 박영희. (2010). 국민건강영양조사 대상자들의 민간의료보험 가입 요인 및 가입여부에 따른 건강행태·의료이용 비교. *한국콘텐츠학회논문지*, 10(12), 190-204.
- 이은현, 김춘자, 조수연, 채현주, 이선희, & 김은정. (2011). 국내 연구에서 당뇨병 환자 건강관련 삶의 질 측정에 관한 모니터링. *Journal of*

Korean Academy of Nursing, 41(4), 558-567.

- 이현옥, & 김진. (2008). 노인의 구강건강신념과 구강건강관리행위가 삶의 질에 미치는 영향. *치위생과학회지*, 8(2), 57-63.
- 이현주. (2011). *Random Forest 분류기에 의한 부정맥 데이터 분석에 대한 연구*. 세종대학교 대학원, 서울.
- 이현지. (2007). 만성질환 노인을 돌보는 가족부양자의 건강관련행동변화에 관한 연구. *한국가족복지학*, 19, 29-53.
- 임수종, 민옥기, & Sw원천기술연구팀. (2012). 빅데이터 활용을 위한 기계학습 기술동향. *전자통신동향분석*, 2012(4), 55-63.
- 장남식, 홍성완, & 장재호. (2000). 성공적인 지식경영을 위한 핵심정보기술-데이터마이닝: 대청.
- 장병탁. (2007). 차세대 기계학습 기술. *정보과학회지*, 25(3), 96-107.
- 장선주. (2010). *제2형 당뇨병노인의 건강관련 삶의 질 구조모형*. 서울대학교 대학원, 서울.
- 전은영, & 최연희. (2010). 취약계층 남성노인의 연령별 건강관련 삶의 질과 영향요인. *Journal of Korean Academy of Nursing*, 40(3), 400-410.
- 정승일. (2010). *분류나무, SVM과 로지스틱 회귀모형의 비교*. 서울대학교 대학원, 서울.
- 정영호. (2011). 한국의료패널로 본 우리나라 성인의 건강관련 삶의 질. *보건복지포럼 통권*(182), 2011.2012.
- 정재임, & 김경아. (2002). 만성질환을 가진 노인들의 건강상태와 생활습관에 관한 조사연구. *경북전문대학논문집*, 21, 375-396.
- 정지연, 김준수, 최현정, 이가영, & 박태진. (2009). 한국 노인의 일상생활수행능력 및 도구적 일상생활 수행능력과 관련된 요인: 국민건강영양조사 제3기 (2005년) 자료를 중심으로. *Korean Journal of Family Medicine*, 30, 598-609.
- 조은주. (2009). 노인여가활동과 라이프스타일 및 삶의 질의 관계. *미간행*

박사학위논문, 부산교육대학교 대학원.

질병관리본부. (2011). 2010년 국민건강영양조사 결과

차남현. (2010). 일 지역사회 노인의 양생에 대한 연구. *지역사회간호학회지*, 21(3), 293-302.

차은정. (2012). 기러기 아빠의 건강관련 삶의 질 예측모형 구축. *성인간호학회지*, 24(4), 428-437.

최동운, & 김수용. (2010). 스마트폰 기반의 생체인식 데이터 마이닝 알고리즘. *한국엔터테인먼트산업학회 학술대회 논문집*, 7(2), 155-158.

최령, 문현주, & 함병덕. (2010). 만성질환이 노인들의 스트레스인지, 우울 증상 경험 및 자살생각에 미치는 영향. *보건의료산업학회지*, 4(2), 73-84.

최선영. (2009). *데이터마이닝을 이용한 건강관련 삶의 질의 예측 요인*. 고려대학교 보건대학원, 서울.

최수정. (2001). *도구개발을 통한 한국노인의 삶의 질에 관한 조사연구*. 서울.

최연희. (2004). 농촌지역 독거노인의 건강증진행위와 삶의 질. *보건교육·건강증진학회지*, 21(2), 87-100.

최영희. (2007). *노인과 건강*. 서울: 현문사.

최윤진. (2009). *연령에 따른 노인의 삶의 질에 영향을 미치는 요인에 관한 연구*. 한영신학대학교 대학원.

최정숙, 권성옥, & 백희영. (2004). 농촌지역 노인들의 사회경제적인 요인 및 건강습관에 따른 건강관련 삶의 질. *농촌의학.지역보건*, 29(1), 29-41.

최해림. (2008). *고혈압, 당뇨병과 건강 관련 삶의 질의 관련성*. 연세대학교 대학원, 서울.

하현근. (2008). *뇌졸중 관련 특성에 따른 삶의 질*. 연세대학교 보건대학원, 서울.

- 한국보건사회연구원. (2012). *2011년도 노인실태조사*.
- 한미아, 류소연, 박종, 강명근, 박종구, & 김기순. (2008). 일부 농촌지역 성
인에서 EQ-5D 를 이용한 건강관련 삶의 질. *예방의학회지*, 41(3),
173-180.
- 한은정. (2005). *건강검진 자료에서 Random forests를 이용한 백내장 발
생 위험군 예측모형*. 연세대학교 대학원, 서울.
- 홍인옥. (2011). *흡연, 음주, 운동, 비만 및 인구사회학적 특성에 따른 건강
관련 삶의 질*. 삼육대학교 보건대학원, 서울.

Appendix

Appendix 1. Survey of KNHANES

KNHANES의 추진 경과를 살펴보면, 1969년에 도입된 국민영양조사와 1971년에 도입된 국민건강 및 보건의식행태조사를 통합하여 1998년부터 시작되었으며, 1995년에 공포된 국민건강증진법 제16조에 의거하여 1998년부터 3년주기로 시행되는 전국규모의 건강 및 영양조사이다(KCDC, 2010) (그림 28).

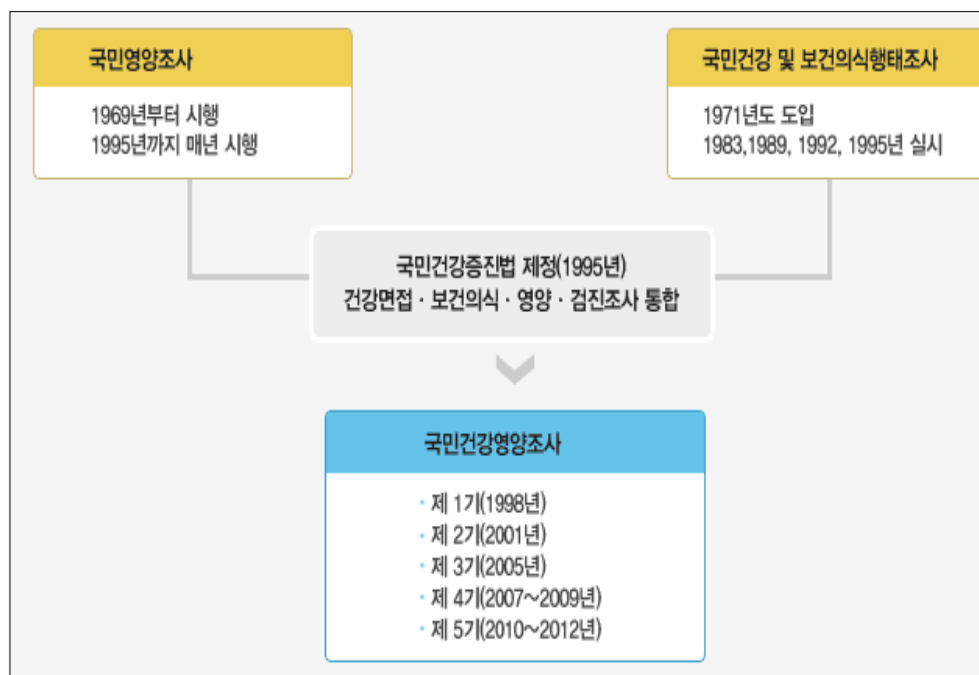


Figure 28. History of KNHANES implementation

KNHANES 원시자료는 홈페이지(<http://knhanes.cdc.go.kr>)를 통해 무료로 다운로드 받을 수 있다

KNHANES는 제1~3기까지는 3년의 간격을 두고 당해 연도에 2~3개월 동안 실시된 단기조사로 운영하다가, 제4기 1차년도(2007)부터 질병관리본부에서 ‘전문조사 수행팀’을 구성하여 계절적 편향 없이 매년 통계생산이 가능한 연중 조사로 수행하고 있다. 그 자세한 추진 경과는 표 24와 같다(KCDC, 2012).

Table 24. Survey progress of KNHANES

	제1기 (‘98년)	제2기 (‘01년)	제3기 (‘05년)	제4기 (‘07~‘09년)	제5기 (‘10~‘12년)
주기	3년 국가통계	3년 국가통계	3년 국가통계	1년 국가통계 2~3년 시도통계	1년 국가통계 2~3년 시도통계
규모 (검진기준)	200조사구 22~24가구/조사구	200조사구 22가구/조사구	200조사구 22~26가구/조사구	500조사구 23가구/조사구	576조사구 20가구/조사구
기간	11~12월	11~12월	4~6월	연중조사(연50주)	연중조사(연48주)
수행기관	한국보건사회연구원 한국보건산업진흥원	한국보건사회연구원 한국보건산업진흥원	한국보건사회연구원 한국보건산업진흥원 질병관리본부	질병관리본부	질병관리본부

매년 전국 192개 조사구, 3,840가구의 만 1세 이상 가구원 약 1만명을 대상으로 전문조사 수행팀이 매주 4개 지역을 조사하여 연간 48주 조사를 시행하며, 검진조사 및 건강설문조사, 영양조사를 통해 총 525항목을 조사한다. 그 구체적인 조사 내용과 방법은 아래와 같다(질병관리본부, 2011)(표 25, 그림 29).

Table 25. Survey contents of KNHANES

구분	영역	세부내용	항목
계	36	-	525
검진	14	비만, 고혈압, 당뇨병, 이상지혈증, 간질환, 신장질환, 빈혈, 폐질환, 구강질환, 안질환, 이비인후과질환, 골다공증, 골관절염, 중금속	95
건강설문	14	질병이환, 의료이용, 활동제한 및 삶의 질, 손상(사고중독), 예방접종 및 건강검진, 교육 및 경제활동, 흡연, 음주, 신체활동, 비만 및 체중조절, 정신건강, 안전의식, 여성건강, 가구조사	310
영양	8	음식 및 식품섭취내용, 식생활행태, 영양지식, 식이보충제, 식품안전성, 수유현황, 이유보충식, 주요식품별 섭취빈도	120



Figure 29. Survey methods of KNHANES

Appendix 2. Survey table of EQ-5D

다음은 EuroQol Group의 승인을 받아 조사된 EQ-5D 조사표이다.

아래의 각 문항에서 오늘 귀하의 건강상태를 가장 잘 설명해주는 하나의 항목에 표시해 주십시오.

1. 운동능력

1. 나는 걷는데 지장이 없다
 2. 나는 걷는데 다소 지장이 있다
 3. 나는 종일 누워있어야 한다
-

2. 자기 관리

1. 나는 목욕을 하거나 옷을 입는데 지장이 없다
 2. 나는 혼자 목욕을 하거나 옷을 입는데 다소 지장이 있다
 3. 나는 혼자 목욕을 하거나 옷을 입을 수가 없다
-

3. 일상 활동 (예. 일, 공부, 가사일, 가족 또는 여가 활동)

1. 나는 일상 활동을 하는데 지장이 없다
 2. 나는 일상 활동을 하는데 다소 지장이 있다
 3. 나는 일상 활동을 할 수가 없다
-

4. 통증/불편

1. 나는 통증이나 불편감이 없다
 2. 나는 다소 통증이나 불편감이 있다
 3. 나는 매우 심한 통증이나 불편감이 있다
-

5. 불안/우울

1. 나는 불안하거나 우울하지 않다
 2. 나는 다소 불안하거나 우울하다
 3. 나는 매우 심하게 불안하거나 우울하다
-

Appendix 3. Weighted estimate of EQ-5D 243 cases

Health status	EQ-5D Weight	Health status	EQ-5D Weight	Health status	EQ-5D Weight
11111	1.000	21112	0.811	31113	0.324
11112	0.907	21113	0.646	31121	0.445
11113	0.742	21121	0.817	31122	0.402
11121	0.913	21122	0.774	31123	0.287
11122	0.870	21123	0.609	31131	0.331
11123	0.705	21131	0.653	31132	0.288
11131	0.749	21132	0.610	31133	0.173
11132	0.706	21133	0.495	31211	0.431
11133	0.591	21211	0.803	31212	0.388
11211	0.899	21212	0.760	31213	0.273
11212	0.856	21213	0.595	31221	0.394
11213	0.691	21221	0.766	31222	0.351
11221	0.862	21222	0.723	31223	0.236
11222	0.819	21223	0.558	31231	0.280
11223	0.654	21231	0.602	31232	0.237
11231	0.698	21232	0.559	31233	0.122
11232	0.655	21233	0.444	31311	0.274
11233	0.540	21311	0.596	31312	0.231
11311	0.692	21312	0.553	31313	0.116
11312	0.649	21313	0.438	31321	0.237
11313	0.534	21321	0.559	31321	0.237
11321	0.655	21322	0.516	31322	0.194
11322	0.612	21323	0.401	31323	0.079
11323	0.497	21331	0.445	31331	0.123
11331	0.541	21332	0.402	31332	0.080
11332	0.498	21333	0.287	31333	-0.035

Health status	EQ-5D Weight	Health status	EQ-5D Weight	Health status	EQ-5D Weight
11333	0.383	22111	0.808	32111	0.436
12111	0.904	22112	0.765	32112	0.393
12112	0.861	22113	0.600	32113	0.278
12113	0.696	22121	0.771	32121	0.399
12121	0.867	22122	0.728	32122	0.356
12122	0.824	22123	0.563	32123	0.241
12123	0.659	22131	0.607	32131	0.285
12131	0.703	22132	0.564	32132	0.242
12132	0.660	22133	0.449	32133	0.127
12133	0.545	22211	0.757	32211	0.385
12211	0.853	22212	0.714	32212	0.342
12212	0.810	22213	0.549	32213	0.227
12213	0.645	22221	0.720	32221	0.348
12221	0.816	22222	0.677	32222	0.305
12222	0.773	22223	0.512	32223	0.190
12223	0.608	22231	0.556	32231	0.234
12231	0.652	22232	0.513	32232	0.191
12232	0.609	22233	0.398	32233	0.076
12233	0.494	22311	0.550	32311	0.228
12311	0.646	22312	0.507	32312	0.185
12312	0.603	22313	0.392	32313	0.070
12313	0.488	22321	0.513	32321	0.191
12321	0.609	22322	0.470	32322	0.148
12322	0.566	22323	0.355	32323	0.033
12323	0.451	22331	0.399	32331	0.077
12331	0.495	22332	0.356	32332	0.034
12332	0.452	22333	0.241	32333	-0.081
12333	0.337	23111	0.668	33111	0.346

Health status	EQ-5D Weight	Health status	EQ-5D Weight	Health status	EQ-5D Weight
13111	0.764	23112	0.625	33112	0.303
13112	0.721	23113	0.510	33113	0.188
13113	0.606	23121	0.631	33121	0.309
13121	0.727	23122	0.588	33122	0.266
13122	0.684	23123	0.473	33123	0.151
13123	0.569	23131	0.517	33131	0.195
13131	0.613	23132	0.474	33132	0.152
13132	0.570	23133	0.359	33133	0.037
13133	0.455	23211	0.617	33211	0.295
13211	0.713	23212	0.574	33212	0.252
13212	0.670	23213	0.459	33213	0.137
13213	0.555	23221	0.580	33221	0.258
13221	0.676	23222	0.537	33222	0.215
13222	0.633	23223	0.422	33222	0.215
13223	0.518	23231	0.466	33223	0.100
13231	0.562	23232	0.423	33231	0.144
13232	0.519	23233	0.308	33232	0.101
13233	0.404	23311	0.460	33233	-0.014
13311	0.556	23312	0.417	33311	0.138
13312	0.513	23313	0.302	33312	0.095
13313	0.398	23321	0.423	33313	-0.020
13321	0.519	23322	0.380	33321	0.101
13322	0.476	23323	0.265	33322	0.058
13323	0.361	23331	0.309	33323	-0.057
13331	0.405	23332	0.266	33331	-0.013
13332	0.362	23333	0.151	33332	-0.056
13333	0.247	31111	0.482	33333	-0.171
21111	0.854	31112	0.439		

Appendix 4. Method table of variables normalization

변수명	변수명	상세	정규화
ID_1	개인아이디	조사구번호 거처·주거번호, 가구원 번호	Real Data
ID_fam_1	가구아이디	조사구번호 거처·주거번호	Real Data
year	조사연도	조사연도 4 자리수	Real Data
EQ5D	EQ-5D index	EQ-5D 건강상태에 건강상태 질가 중치 (질병관리본부, 2007) 반영	Real Data
EQ5D_CD_2G	EQ-5D Group	1.>=0.678 0.<0.677	1-of-K encoding
CD_N	만성질환개수	□□개 23 개 상병중 해당 개수 (유병&진단&1 년이상)	Real Data 신규 생성
CD_D	만성질환합산기간	□□□년 23 개 상병 진단후 소요기간 총합 (만연령-진단시기)	Real Data 신규 생성
CD_DM	만성질환평균기간	□□□년 23 개 상병중 유병개수당 평균 소요 기간 (만성질환합산기간/만성질환개수)	Real Data 신규 생성
age	만연령	□□□ 세	Real Data
HE_ht_cg	신장변화량	□□.□ cm DX_Q_ht(젊은시절키)-HE_ht(신 장)	Real Data 신규 생성
DX_F_Ts_A	대퇴골 전체 T-score(아시아 기준)	□.□	Real Data
DX_S_Ts_A	요추 T-score(아시아기준)	□.□	Real Data
DX_FN_Ts_A	대퇴골 경부 T-score(아시아 기준)	□.□	Real Data
cfam	가구원수	□□ 명	Real Data
ainc	월평균 가구총소득	□□□,□□□만원	Real Data
D_2_wk	최근 2 주간 불편감일수	□□ 일 88.비해당(문항 2-②) 99.모름	Real Data 88->0 99->삭제
DI1_ag	고혈압 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DI1_ag_d	고혈압 소요기간	□□년	Real Data 신규 생성
DI2_ag	고지혈증 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DI2_ag_d	고지혈증 소요기간	□□년	Real Data 신규 생성

변수명	변수명	상세	정규화
DI3_ag	뇌졸중 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DI3_ag_d	뇌졸중 소요기간	□□년	Real Data 신규 생성
DI5_ag	심근경색 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DI5_ag_d	심근경색 소요기간	□□년	Real Data 신규 생성
DI6_ag	협심증 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DI6_ag_d	협심증 소요기간	□□년	Real Data 신규 생성
DM2_ag	골관절염 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DM2_ag_d	골관절염 소요기간	□□년	Real Data 신규 생성
DM3_ag	류마티스성 관절염 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DM3_ag_d	류마티스성 관절염 소요기간	□□년	Real Data 신규 생성
DJ2_ag	폐결핵 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DJ2_ag_d	폐결핵 소요기간	□□년	Real Data
DJ4_ag	천식 진단시기(소아,청소년 포함)	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DJ4_ag_d	천식 소요기간	□□년	Real Data 신규 생성
DE1_ag	당뇨병 진단시기(소아,청소년 포함)	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DE1_ag_d	당뇨병 소요기간	□□년	Real Data 신규 생성
DE2_ag	갑상샘 질환 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DE2_ag_d	갑상샘 질환 소요기간	□□년	Real Data 신규 생성

변수명	변수명	상세	정규화
DC1_ag	위암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC1_ag_d	위암 소요기간	□□년	Real Data 신규 생성
DC2_ag	간암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC2_ag_d	간암 소요기간	□□년	Real Data 신규 생성
DC3_ag	대장암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC3_ag_d	대장암 소요기간	□□년	Real Data 신규 생성
DC4_ag	유방암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC4_ag_d	유방암 소요기간	□□년	Real Data 신규 생성
DC5_ag	자궁경부암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC5_ag_d	자궁경부암 소요기간	□□년	Real Data 신규 생성
DC6_ag	폐암 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC6_ag_d	폐암 소요기간	□□년	Real Data 신규 생성
DC11_ag	기타암 1 진단받은 시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DC11_ag_d	기타암 1 소요기간	□□년	Real Data 신규 생성
DF2_ag	우울증 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DL1_ag	아토피 피부염 진단시기 (소아, 청소년 포함)	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DN1_ag	신부전 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제

변수명	변수명	상세	정규화
DN1_ag_d	신부전 소요기간	□□년	Real Data 신규 생성
DK8_ag	B 형간염 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DK8_ag_d	B 형간염 소요기간	□□년	Real Data 신규 생성
DK9_ag	C 형간염 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DK9_ag_d	C 형간염 소요기간	□□년	Real Data 신규 생성
DK4_ag	간경변증 진단시기	□□□ 만_세 888.비해당 999.모름	Real Data 888->0 999->삭제
DK4_ag_d	간경변증 소요기간	□□년	Real Data 신규 생성
LQ2_mn	최근 1 달간 결근일수	□□ 일 88.비해당(문항 3-②⑨) 99.모름	Real Data 88->0 99->삭제
BD2	(만 12 세이상) 음주 시작 연령	만 □□세 88.비해당(문항 4-①⑨) 99.모름	Real Data 88->0 99->삭제
BP8	하루 평균 수면시간	1~24.□□시간 99.모름	Real Data 99->삭제
BS2_1	(성인) 흡연 시작연령	만 □□세 88.비해당(문항 17-③⑨) 99.모름	Real Data 88->0 99->삭제
BS3_2	(성인) 현재흡연자 하루 평 균 흡연량	□□□ 개비 888.비해당 999.모름	Real Data 888->0 999->삭제
BS6_2	(성인) 과거흡연자 흡연기 간 ※월로 환산	□□년 □□개월 8888.비해당 9999.모름	Real Data 8888->0 9999->삭제
BS6_3	(성인) 과거흡연자 하루 평 균 흡연량	□□□ 개비 888.비해당 999.모름	Real Data 888->0 999->삭제
BE3_12_13	격렬한 신체활동 지속 시간 ※분으로 환산	□□□분 88.전혀 하지 않음 99.모름	Real Data 88->0 99->삭제
BE3_22_23	중등도 신체활동 지속 시간 ※분으로 환산	□□□분 88.전혀 하지 않음 99.모름	Real Data 88->0 99->삭제

변수명	변수명	상세	정규화
BE3_32_33	걷기 지속 시간(시간)	□□□분 88.전혀 하지 않음 99.모름	Real Data 88->0 99->삭제
HE_fst	공복시간	□□ 시간	Real Data
HE_sbp_tr	최종 수축기혈압(보정값)	□□□ mmHg	Real Data
HE_dbp_tr	최종 이완기혈압(보정값)	□□□ mmHg	Real Data
HE_wc	허리둘레	□□□.□ cm	Real Data
HE_BMI	체질량지수	□□□.□ kg/m	Real Data
HE_glu	공복혈당	□□.□ mg/dl	Real Data
HE_insulin	인슐린	□□□□.□□ μIU/mL	Real Data
O_DTD	우식유치수	□□ 개	Real Data
O_DTP	우식영구치수	□□ 개	Real Data
O_DFTD	우식경험 유치수	□□ 개	Real Data
O_DMFTP	우식경험 영구치수	□□ 개	Real Data
N_WAT_C	물섭취량(컵)	물섭취량(컵:200ml)	Real Data
AC3	손상 횟수	□□번 88.비해당(문항 1-②⑨) 99.모름	Real Data 88->0 99->삭제
AC8_2_01_03	손상 1~3 와병 합산일수	□□ 일 (손상 1 + 손상 3 와병일수)	Real Data 신규 생성
AC8_2_01~03	손상 1~3 와병일수	□□ 일 88.비해당(문항 6-②⑨) 99.모름	Real Data 88->0 99->삭제
MH1_1	입원이용 횟수(0~7)	□□번 88.비해당(문항 1-②⑨) 99.모름	Real Data 88->0 99->삭제
MH5_1_01_06	입원 1~6 합산 입원기간	□□□일 (일로 환산)	Real Data 신규 생성
MH5_1_01~06	입원 1~6 입원기간	□□□일 (일로 환산) 999.모름	Real Data 999->삭제
MH7_1_01_06	입원 1~6 합산 본인부담금	□□,□□□,□□□ 원	Real Data 신규 생성
MH7_1_01~6	입원 1~6 본인부담금	□□,□□□,□□□ 원 99999999.모름	Real Data 99999999->삭제
MH7_2	입원 1~6 일단위 본인부담금	□□,□□□,□□□ 원 입원 1~6 합산 본인부담금/입원 1~6 합산 입원기간	Real Data 신규 생성
MO1_1	외래이용 횟수	□□번 88.비해당(문항 1-②⑨) 99.모름	Real Data 88->0 99->삭제
MO7_1_01_14	외래 1~14 합산 본인부담금	□□,□□□,□□□ 원	Real Data 신규 생성

변수명	변수명	상세	정규화
MO7_1_01~14	외래 1~14 본인부담금	□□,□□□,□□□ 원 99999999.모름	Real Data 99999999->삭제
MO7_2	외래 1~14 건당 본인부담금	□□,□□□,□□□ 원 외래 1~14 합산 본인부담금/외래이용 횟수	Real Data 신규 생성
MP1_1	약국이용횟수	□□번 88.비해당(문항 1-②⑨) 99.모름	Real Data 88->0 99->삭제
LQ1_mn	최근 1 달간 외병일수	□□ 일 88.비해당(문항 2-②⑨) 99.모름	Real Data 88->0 99->삭제
MH4_1_01~06	입원 1~6 이용이유 -질병 코드(1-230)	□□□ 999.모름	1-of-K encoding 999->삭제
MO4_1_01~14	외래 1~14 이용이유 -질병 코드	□□□ 999.모름	1-of-K encoding 999->삭제
town_t	동읍면	1.동 2.읍·면	1-of-K encoding
apt_t	주택유형	1.일반 2.아파트	1-of-K encoding
sex	성별	1.남자 2.여자	1-of-K encoding
incm	소득 4 분위수(개인)	1.하 2.중하 3.중상 4.상	1-of-K encoding
ho_incmm	소득 4 분위수(가구)	1.하 2.중하 3.중상 4.상	1-of-K encoding
edu	교육수준 재분류 코드	1.초졸이하 2.중졸 3.고졸 4.대졸이상	1-of-K encoding
occp	직업재분류 및 실업/비경제 활동인구상태 코드	1.관리자,전문가 및 관련 종사자 2.사무종사자 3.서비스 및 판매 종사자 4.농림어업 숙련 종사자 5.기능원,장치,기계조작및조립종사자 6.단순노무종사자 7.무직(주부,학생 등)	1-of-K encoding
DX_Q_hsty	골다공증가족력	1.예 2.아니오 9.무응답	1-of-K encoding
DX_Q_MP	폐경여부(최종)	1.예 2.아니오 3.양측난소절제술 8.비해당(남자) 9.무응답	1-of-K encoding

변수명	변수명	상세	정규화
DX_OST	골다공증유병여부	1.정상 2.골감소증 3.골다공증	1-of-K encoding
DX_OST_TF	골다공증유병여부(대퇴골 전체)	1.정상 2.골감소증 3.골다공증	1-of-K encoding
DX_OST_FN	골다공증유병여부(대퇴골 경부)	1.정상 2.골감소증 3.골다공증	1-of-K encoding
DX_OST_LS	골다공증유병여부(요추)	1.정상 2.골감소증 3.골다공증	1-of-K encoding
genertn	가구세대구성코드	10.1 인 가구 11.부부 (응답자+ 배우자) 12.부부+ 미혼 형제자매 13.응답자+ 미혼 형제자매 14.부부+ 기타 친인척 15.응답자+ 기타 친인척 19.상기를 제외한 모든 1 세대 가구 21.부부+ 미혼자녀 22.편부+ 미혼자녀 23.편모+ 미혼자녀 24.부부+ 양친 25.부부+ 편부모 26.부부+ 자녀+ 부부의 형제또는자매 27.조부모+ 미혼 손자녀 28.편조부/편조모+ 미혼손자녀 29.상기를 제외한 모든 2 세대가구 31.부부+ 미혼자녀+ 양친 32.부부+ 미혼자녀+ 편부모 39.상기를 제외한 모든 3 세대이상	1-of-K encoding
allownc	기초생활수급 여부	1.그렇다 2.지금은 아니다 3.아니다 9.모름	1-of-K encoding
house	주택소유여부	1.아니오 2.1 채 있음 3.2 채 이상 있음 9.모름	1-of-K encoding
live_t	주택형태	1.단독주택 2.아파트 3.연립주택 4.다세대 주택 5.영업용 건물 내 주택 6.기타 9.모름	1-of-K encoding
ainc_unit1	가구총소득(개방형) : 소득 단위	1.연 2.월 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
marri_1	결혼여부	1.기혼 2.미혼 9.모름	1-of-K encoding
marri_2	결혼상태	1.유배우자,동거 2.유배우자,별거 3.사별 4.이혼 8.비해당(문항 10-②) 9.모름	1-of-K encoding
tins	건강보험종류	1.지역의료보험 2.사업장(직장)가입자 3.의료급여 1 종 4.의료급여 2 종 5.미가입 9.모름	1-of-K encoding
npins	민간의료보험가입여부	1.예 2.아니오 9.모름	1-of-K encoding
D_1_1	주관적 건강상태	1.매우 좋음 2.좋음 3.보통 4.나쁨 5.매우 나쁨 9.모름	1-of-K encoding
D_2_1	최근 2 주간 몸이 불편했던 경험 유무	1.예 2.아니오 9.모름	1-of-K encoding
DI1_lt	고혈압 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DI1_dg	고혈압 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI1_pr	고혈압 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI1_pt	고혈압 치료	1.있음 0.없음 8.비해당(지금까지 앓은 적 없음)	1-of-K encoding
DI1_2	혈압조절제 복용	1.매일 복용함 2.한달에 20 일 이상 복용함 3.한달에 15 일 이상 복용함 4.한달에 15 일 미만 복용함 5.복용하지 않음 8.비해당(현재 앓고 있지 않음) 9.모름	1-of-K encoding
DI1_3	고혈압관리교육	1.예 2.아니오 8.비해당(현재 앓고 있지 않음) 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
DI2_lt	고지혈증 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DI2_dg	고지혈증 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI2_pr	고지혈증 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI2_pt	고지혈증 치료	1.있음 0.없음 8.비해당(지금까지 앓은 적 없음)	1-of-K encoding
DI2_2	고지혈증제 복용	1.매일 복용함 2.한달에 20 일 이상 복용함 3.한달에 15 일 이상 복용함 4.한달에 15 일 미만 복용함 5.복용하지 않음 8.비해당 9.모름	1-of-K encoding
DI3_lt	뇌졸중 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DI3_dg	뇌졸중 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI3_pr	뇌졸중 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI3_pt	뇌졸중 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DI3_2	뇌졸중 후유증	1.후유증을 앓고 있음 2.후유증이있었지만현재회복되었음 3.특별한 후유증이 없음 8.비해당 9.모름	1-of-K encoding
DI4_lt	심근경색증, 협심증 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI4_dg	심근경색증, 협심증 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI4_pr	심근경색증, 협심증 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI4_pt	심근경색증, 협심증 치료	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DI5_lt	심근경색 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DI5_dg	심근경색 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI5_pr	심근경색 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI5_pt	심근경색 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DI6_lt	협심증 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DI6_dg	협심증 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI6_pr	협심증 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DI6_pt	협심증 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DM1_lt	관절염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DM1_dg	관절염 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM1_pr	관절염 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM1_pt	관절염 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DM2_lt	골관절염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DM2_dg	골관절염 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM2_pr	골관절염 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM2_pt	골관절염 치료	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DM3_lt	류마티스성 관절염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DM3_dg	류마티스성 관절염 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM3_pr	류마티스성 관절염 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DM3_pt	류마티스성 관절염 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DM1_5	관절염관리교육	1.예 2.아니오 8.비해당 9.모름	1-of-K encoding
DJ2_lt	폐결핵 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DJ2_dg	폐결핵 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DJ2_pr	폐결핵 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DJ2_pt	폐결핵 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DJ2_4	결핵 주 치료기관	1.보건소 2.병의원 3.기타	1-of-K encoding
DJ4_lt	천식 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DJ4_dg	천식 의사진단여부 (소아, 청소년 포함)	1.있음 0.없음 8.비해당 9.모름	1-of-K encoding
DJ4_pr	천식 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DJ4_pt	천식 치료	1.있음 0.없음 8.비해당	1-of-K encoding
D_3_1	1년간 숨쉴때 가슴에서 소리 여부	1.예 2.아니오 9.모름	1-of-K encoding
D_3_2	1년간 운동중에 가슴에서 소리 여부	1.예 2.아니오 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
DJ4_3	천식 약복용	1.악화방지 및 천식증산관리를 위해 정기적으로 치료를 받음 2.증상이 있을 때만 치료를 받음 3.치료받지 않음 8.비해당 9.모름	1-of-K encoding
DE1_lt	당뇨병 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DE1_dg	당뇨병 의사진단여부 (소아, 청소년 포함)	1.있음 0.없음 8.비해당 9.모름	1-of-K encoding
DE1_pr	당뇨병 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DE1_pt	당뇨병 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DE1_3	당뇨병치료여부	1.예 2.아니오 8.비해당	1-of-K encoding
DE1_31	당뇨병치료_인슐린 주사	1.예 0.아니오 8.비해당	1-of-K encoding
DE1_32	당뇨병치료_당뇨병약	1.예 0.아니오 8.비해당	1-of-K encoding
DE1_33	당뇨병치료_비약물 요법	1.예 0.아니오 8.비해당	1-of-K encoding
DE1_34	당뇨병치료_기타	1.예 0.아니오 8.비해당	1-of-K encoding
DE1_5	당뇨병관리교육	1.예 2.아니오 8.비해당 9.모름	1-of-K encoding
DE2_lt	갑상샘 질환 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DE2_dg	갑상샘 질환 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DE2_pr	갑상샘 질환 현재유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DE2_pt	갑상샘 질환 치료	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DC1_lt	위암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC1_dg	위암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC1_pr	위암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC1_pt	위암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC2_lt	간암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC2_dg	간암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC2_pr	간암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC2_pt	간암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC3_lt	대장암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC3_dg	대장암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC3_pr	대장암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC3_pt	대장암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC4_lt	유방암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC4_dg	유방암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC4_pr	유방암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC4_pt	유방암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DC5_lt	자궁경부암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC5_dg	자궁경부암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC5_pr	자궁경부암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC5_pt	자궁경부암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC6_lt	폐암 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC6_dg	폐암 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding
DC6_pr	폐암 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC6_pt	폐암 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC11_tp_1	기타암 1 의 종류	1.무응답 2.갑상샘암 4.골수암 5.구강암 6.난소암 7.뇌종양 8.림프암 9.방광암 10.식도암 11.신장암 12.췌장암 13.요로점막의악성신생물 14.자궁암 15.전립샘암 17.직장암 18.췌장암 19.편도암 20.피부암 22.혈액암 23.후두암	1-of-K encoding
DC11_lt	기타암 1 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DC11_dg	기타암 1 의사진단	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DC11_pr	기타암 1 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DC11_pt	기타암 1 치료 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DF2_lt	우울증 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DF2_dg	우울증 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DF2_pr	우울증 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DF2_pt	우울증 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DL1_lt	아토피 피부염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DL1_dg	아토피 피부염 의사진단여부 (소아, 청소년 포함)	1.있음 0.없음 8.비해당	1-of-K encoding
DL1_pr	아토피 피부염 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DL1_pt	아토피 피부염 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DN1_lt	신부전 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DN1_dg	신부전 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DN1_pr	신부전 현재 유병 여부	1.있음 0.없음 8.비해당	1-of-K encoding
DN1_pt	신부전 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DK8_lt	B 형간염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DK8_dg	B 형간염 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
DK8_pr	B 형간염 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DK8_pt	B 형간염 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DK9_lt	C 형간염 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DK9_dg	C 형간염 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DK9_pr	C 형간염 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DK9_pt	C 형간염 치료	1.있음 0.없음 8.비해당	1-of-K encoding
DK4_lt	간경변증 유병여부	1.있음 0.없음 9.무응답	1-of-K encoding
DK4_dg	간경변증 의사진단여부	1.있음 0.없음 8.비해당	1-of-K encoding
DK4_pr	간경변증 현재 유병여부	1.있음 0.없음 8.비해당	1-of-K encoding
DK4_pt	간경변증 치료	1.있음 0.없음 8.비해당	1-of-K encoding
BH9_11	인플루엔자(독감) 예방접종 여부	1.예 2.아니오 9.모름	1-of-K encoding
LQ4_00	활동제한 여부	1.예 2.아니오 9.모름	1-of-K encoding
LQ4_01	활동제한 사유 : 골절, 관절 부상	1.예 0.아니오 8.비해당(문항 1-②⑨)	1-of-K encoding
LQ4_02	활동제한 사유 : 기타 사고	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_03	(성인) 활동제한 사유: 관절 염, 류마티즘	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_04	활동제한 사유 : 심장질환	1.예 0.아니오 8.비해당	1-of-K encoding

변수명	변수명	상세	정규화
LQ4_05	활동제한 사유: 호흡문제, 폐질환, 천식	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_06	활동제한 사유: 뇌졸중	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_07	활동제한 사유: 당뇨병	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_08	활동제한 사유: 고혈압	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_09	활동제한 사유: 등, 목의 문제	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_10	활동제한 사유: 암	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_11	활동제한 사유: 치아 및 구강질환	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_12	활동제한 사유: 시력문제	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_13	활동제한 사유: 청각문제	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_14	활동제한 사유: 치매	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_15	활동제한 사유: 우울/불안/정서상 문제	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_16	활동제한 사유: 정신지체	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_21	활동제한 사유: 비만	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_22	활동제한 사유: 노령	1.예 0.아니오 8.비해당	1-of-K encoding
LQ4_23	활동제한 사유: 기타	1.예 0.아니오 8.비해당	1-of-K encoding
LQ1_sb	최근 1 달간 와병여부	1.예 2.아니오 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
LQ2_ab	최근 1 달간 결근여부	1.예 2.아니오 9.모름	1-of-K encoding
educ	교육수준 - 학력	1.서당/한학 2.무학 3.초등학교 4.중학교 5.고등학교 6.2년/3년제 대학 7.4년제 대학 8.대학원 88.소아 99.모름	1-of-K encoding
graduat	교육수준 - 졸업여부	1.졸업 2.수료 3.중퇴 4.재학/휴학 중 8.비해당 9.모름	1-of-K encoding
EC1_1	경제활동 상태	1.예(취업자) 2.아니오(실업자,비경제활동인구) 9.모름	1-of-K encoding
EC1_2	미취업 사유	1.필요를 느끼지 못해서 2.학교/학원에 재학/재원 중 3.정년/퇴직 상태 4.건강상의 이유 5.실업/구직 중 6.육아/간병 등 7.기타 8.비해당 9.모름	1-of-K encoding
EC_pedu_1	(성인) 유년기환경 : 아 버지 교육수준	1.무학 2.서당/한학 3.초등학교 졸업 이하 4.중학교 졸업 이하 5.고등학교 졸업 이하 6.2년/3년제 대학 졸업 이하 7.4년제 대학 졸업 이하 8.대학원 수료 이상 88.소아청소년 99.모름	1-of-K encoding
EC_pedu_2	(성인) 유년기환경 : 어 머니 교육수준	1.무학 2.서당/한학 3.초등학교 졸업 이하 4.중학교 졸업 이하 5.고등학교 졸업 이하 6.2년/3년제 대학 졸업 이하 7.4년제 대학 졸업 이하 8.대학원 수료 이상 88.소아청소년 99.모름	1-of-K encoding

변수명	변수명	상세	정규화
BO1	주관적 체형인식	1.매우 마른 편 2.약간 마른 편 3.보통 4.약간 비만 5.매우 비만임 9.모름	1-of-K encoding
BO1_1	1 년간 체중 변화 여부	1.변화 없음 2.체중 감소 3.체중 증가 9.모름	1-of-K encoding
BO1_2	(성인) 1 년간 체중 감소량	1.3kg 이상 -6kg 미만 2.6kg 이상 -10kg 미만 3.10kg 이상 8.비해당(문항 2-①③⑨) 9.모름	1-of-K encoding
BO1_3	(성인) 1 년간 체중 증가량	1.3kg 이상 -6kg 미만 2.6kg 이상 -10kg 미만 3.10kg 이상 8.비해당(문항 2-①②⑨) 9.모름	1-of-K encoding
BO2_1	1 년간 체중 조절 여부	1.체중 감소 노력 2.체중 유지 노력 3.체중 증가 노력 4.체중 조절 노력해본 적 없음 9.모름	1-of-K encoding
BD1	(만 12 세이상) 평생음주경 험	1.술을 마셔 본 적 없음 2.있음 9.모름	1-of-K encoding
BD1_11	(만 12 세이상) 1 년간 음주 빈도	1.최근 1 년간 전혀 마시지 않았다 2.월 1 회미만 3.월 1 회정도 4.월 2-4 회 5.주 2-3 회정도 6.주 4 회이상 8.비해당(문항 4-①⑨) 9.모름	1-of-K encoding
BD2_1	(만 12 세이상) 한번에 마 시는 음주량	1.1-2 잔 2.3-4 잔 3.5-6 잔 4.7-9 잔 5.10 잔 이상 8.비해당 9.모름	1-of-K encoding
BD2_31	(성인) 남자 고위험음주 빈 도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
BD2_32	(성인) 여자 고위험음주 빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD4	(성인) 음주 자제 안 된 경우의 빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD5	(성인) 음주로 인한 일상생활 지장 빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD6	(성인) 음주 다음날 아침 해장술 빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD7_1	(성인) 음주 후 후회 빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD7_2	(성인) 음주 후 취중 일 망각빈도	1.전혀 없음 2.월 1 회미만 3.월 1 회정도 4.주 1 회정도 5.거의매일 8.비해당 9.모름	1-of-K encoding
BD7_3	(성인) 음주로 인한 상해여부	1.없었다 2.최근 1 년 동안에는 없었다 3.최근 1 년동안에 그런적이 있었다 8.비해당(문항 4-①⑨) 9.모름	1-of-K encoding

변수명	변수명	상세	정규화
BD7_4	(성인) 가족/의사의 절주 권고 여부	1.없었다 2.최근 1 년 동안에는 없었다 3.최근 1 년동안에 그런적이 있었다 8.비해당(문항 4-①⑨) 9.모름	1-of-K encoding
BD7_5	(성인) 1 년간 음주문제 상담 여부	1.예 2.아니오(문항 4-①⑨) 9.모름	1-of-K encoding
dr_month	월간음주여부	0.평생비음주(문항 4-②),최근 1 년 간 월 1 잔 미만 음주(문항 4-① & 문항 5-1-①②) 1.최근 1 년간 월 1 잔이상 음주	1-of-K encoding
BP1	평소 스트레스 인지 정도	1.대단히 많이 느낀다 2.많이 느끼는 편이다 3.조금 느끼는 편이다 4.거의 느끼지 않는다 9.모름	1-of-K encoding
BP5	2 주이상 연속 우울감 여부	1.예 2.아니오 9.모름	1-of-K encoding
BP6_10	1 년간 자살 생각 여부	1.예 2.아니오 9.모름	1-of-K encoding
BP6_31	1 년간 자살 시도 여부	1.예 2.아니오 8.비해당(문항 16-②) 9.모름	1-of-K encoding
BP7	1 년간 정신문제 상담	1.예 2.아니오 9.모름	1-of-K encoding
mh_stress	스트레스인지여부	1.스트레스 많이 느낌 0.스트레스 적게 느낌	1-of-K encoding
mh_melan	우울증상경험여부	1.2 주이상 연속 우울감 0.2 주이상 연속 우울감 없음	1-of-K encoding
mh_suicide	자살생각여부	1.자살 생각해본 적 있음 0.자살 생각해본 적 없음	1-of-K encoding
BS1_1	(성인) 평생흡연 여부	1.5 갑(100 개비)미만 2.5 갑(100 개비)이상 3.피운 적 없음 9.모름	1-of-K encoding
BS3_1	(성인) 현재흡연 여부	1.피움 2.가끔피움 3.과거엔 피웠으나,현재피우지않음 8.비해당(문항 17-③⑨) 9.모름	1-of-K encoding
sm_presnt	현재흡연여부	1.현재흡연 0.과거 흡연,비흡연	1-of-K encoding

변수명	변수명	상세	정규화
BE3_11	1 주일간 격렬한 신체활동 일수	1.전혀하지않음 2.1 일 3.2 일 4.3 일 5.4 일 6.5 일 7.6 일 8.7 일(매일) 88.비해당 99.모름	1-of-K encoding
BE3_21	1 주일간 중등도 신체활동 일수	1.전혀하지않음 2.1 일 3.2 일 4.3 일 5.4 일 6.5 일 7.6 일 8.7 일(매일) 88.비해당 99.모름	1-of-K encoding
BE3_31	1 주일간 걷기 일수	1.전혀하지않음 2.1 일 3.2 일 4.3 일 5.4 일 6.5 일 7.6 일 8.7 일(매일) 88.비해당 99.모름	1-of-K encoding
BE5_1	1 주일간 근력운동 일수	1.전혀하지않음 2.1 일 3.2 일 4.3 일 5.4 일 6.5 일이상 8.비해당 9.모름	1-of-K encoding
BE5_2	1 주일간 유연성운동 일수	1.전혀하지않음 2.1 일 3.2 일 4.3 일 5.4 일 6.5 일이상 8.비해당 9.모름	1-of-K encoding
pa_high	격렬한신체활동실천여부	1.격렬한신체활동 1 회 20 분이상, 주 3 일이상 실천 0.격렬한신체활동 1 회 20 분이상, 주 3 일이상 실천하지 않음	1-of-K encoding

변수명	변수명	상세	정규화
pa_mid	중등도신체활동실천여부	1.중등도 신체활동 1 회 30 분이상, 주 5 일이상 실천 0.중등도 신체활동 1 회 30 분이상, 주 5 일이상 실천하지 않음	1-of-K encoding
pa_walk	걷기실천여부	1.걷기 1 회 30 분이상,주 5 일이상 실천 0.걷기 1 회 30 분이상,주 5 일이상 실천하지 않음	1-of-K encoding
HE_ALC	검진당일 알코올섭취여부	0.아니오 1.예	1-of-K encoding
HE_HPdg	고혈압 의사진단 여부	0.아니오 1.예	1-of-K encoding
HE_HPdr	검진당일 고혈압 약복용 여 부	0.아니오 1.예	1-of-K encoding
HE_DMdg	당뇨병 의사진단 여부	0.아니오 1.예	1-of-K encoding
HE_DMdr	검진당일 당뇨병 약복용 여 부	0.아니오 1.예	1-of-K encoding
HE_rPLS	맥박규칙성 여부	1.규칙적 2.불규칙적	1-of-K encoding
HE_HP_tr	고혈압 유병여부 (보정값 30 세이상)	1.정상 2.고혈압전단계 3.고혈압	1-of-K encoding
HE_obe	비만 유병여부(19 세이상)	1.저체중 2.정상 3.비만	1-of-K encoding
HE_DM	당뇨병 유병여부(30 세이 상)	1.정상 2.공복혈당장애 3.당뇨병	1-of-K encoding
HE_hCHOL	고콜레스테롤혈증 유병여 부(30 세이상)	0.없음 1.있음	1-of-K encoding
HE_IHDL_tr	저 HDL 콜레스테롤혈증 유 병여부(30 세이상)	0.없음 1.있음	1-of-K encoding
HE_hepaB	B 형간염표면항원 양성여 부(10 세이상)	0.음성 1.양성	1-of-K encoding
HE_anem	빈혈 유병여부(10 세이상)	0.없음 1.있음	1-of-K encoding
O_DID	유치 우식유병여부	0.우식유치 수 =0 1.우식유치 수 ≥ 1	1-of-K encoding
O_DIP	영구치 우식유병여부	0.우식영구치 수 =0 1.우식영구치 수 ≥ 1	1-of-K encoding
O_DFID	유치 우식경험여부	0.우식경험유치 수 =0 1.우식경험유치 수 ≥ 1	1-of-K encoding
O_DMFIG	영구치 우식경험여부	0.우식경험영구치 수 =0 1.우식경험영구치 수 ≥ 1	1-of-K encoding

변수명	변수명	상세	정규화
OR1	본인인지 구강건강상태	1.매우 좋음 2.좋음 3.보통 4.나쁨 5.매우 나쁨	1-of-K encoding
BM7	(성인) 씹기문제	1.매우 불편함 2.불편함 3.그저 그러함 4.불편하지 않음 5.전혀 불편하지 않음 8.비해당(19 세미만)	1-of-K encoding
BM8	(성인) 말하기문제	1.매우 불편함 2.불편함 3.그저 그러함 4.불편하지 않음 5.전혀 불편하지 않음 8.비해당(19 세미만)	1-of-K encoding
N_DIET	식이요법여부	1.예 2.아니오 9.모름/무응답	1-of-K encoding
N_DIET_WHY	식이요법 상세이유	1.질환이 있어서 2.체중을 조절하기 위해서 3.기타 8.비해당(식이요법여부:아니오) 9.모름/무응답	1-of-K encoding
T_Q_DZ	설문 2:어지럼증,균형감이상경험	1.예 2.아니오 8.비해당(40 세미만) 9.무응답	1-of-K encoding
T_Q_DZ1	설문 2-1:심한어지러움경험	1.현재는 그렇다 2.1 년이내 그런적이 있다 3.없다 8.비해당 9.무응답	1-of-K encoding
T_Q_DZ2	설문 2-2:넘어짐,쓰러짐여부	1.예 2.아니오 8.비해당(40 세미만) 9.무응답	1-of-K encoding
T_Q_VC	설문 3:본인 목소리 이상인지	1.예 2.아니오 8.비해당(19 세미만) 9.무응답	1-of-K encoding
T_Q_VC1	설문 3-1:목소리이상발생기간	1.3 주이내 2.3 주이상 8.비해당(19 세미만) 9.무응답	1-of-K encoding

변수명	변수명	상세	정규화
T_Q_HR	설문 4:본인청력	1.불편하지 않다 2.약간 불편하다 3.많이 불편하다 4.전혀 들리지 않다 8.비해당(12 세미만) 9.무응답	1-of-K encoding
T_Q_HR1	설문 4-1:보청기,인공와우 사용여부	1.예 2.있지만 거의 사용 안한다 3.아니오 8.비해당(12 세미만) 9.무응답	1-of-K encoding
T_Q_VN	설문 5:이명여부	1.예 2.아니오 3.기억이 나지 않는다 8.비해당(12 세미만) 9.무응답	1-of-K encoding
T_Q_VN1	설문 5-1:이명에따른 생활 불편	1.불편하지 않다 2.성가시고 신경이 쓰인다 3.잠을 이루기가 힘들 정도이다 8.비해당(12 세미만) 9.무응답	1-of-K encoding
T_Q_CR	설문 6:비염증상 경험	1.예 2.아니오 8.비해당(6 세미만) 9.무응답	1-of-K encoding
T_Q_CR1	설문 6-1: 발생 계절	1.특정계절에만 나타난다 2.일년 내내 나타난다 8.비해당(6 세미만) 9.무응답	1-of-K encoding
T_Q_CR2	설문 6-2:일주일 내 지속기 간	1.4 일 미만 2.4 일 이상 8.비해당 9.무응답	1-of-K encoding
T_Q_CR3	설문 6-3:지속기간	1.1 달미만 2.1 달이상 8.비해당 9.무응답	1-of-K encoding
T_Q_CR4	설문 6-4:공부,일,수면 방 해여부	1.예 2.아니오 8.비해당 9.무응답	1-of-K encoding
T_Q_SNST1	설문 7-1:비루,후비루 3 개 월이상 지속여부	1.없다 2.있다 8.비해당(6 세미만) 9.무응답	1-of-K encoding
T_Q_SNST2	설문 7-2:코막힘 3 개월이 상 지속여부	1.없다 2.있다 8.비해당(6 세미만) 9.무응답	1-of-K encoding
T_Q_SNST3	설문 7-3:안면부동통,압박 감 3 개월이상 지속여부	1.없다 2.있다 8.비해당(6 세미만) 9.무응답	1-of-K encoding

변수명	변수명	상세	정규화
T_Q_SNST4	설문 7-4:후각장애 3 개월 이상 지속여부	1.없다 2.있다 8.비해당(6 세미만) 9.무응답	1-of-K encoding
T_NC_anism	후각장애증상유병여부	0.없음 1.있음	1-of-K encoding
T_VCds	본인인지음성장애유병여부	0.없음 1.있음	1-of-K encoding
T_sAlgrn	비염유병여부	0.없음 1.있음	1-of-K encoding
T_Mscp	후두내시경	0.정상 1.비정상 2.관정불가 8.비해당(19 세미만) 9.무응답(미검사)	1-of-K encoding
T_SNST	만성비루비동염유병여부	0.없음 1.있음	1-of-K encoding
T_Nc_dev	비중격만곡증유병여부	0.없음 1.있음	1-of-K encoding
E_Q_FAM	설문 3:안과 질환 가족력여부	1.예 2.아니오 8.비해당(12 세미만) 9.무응답	1-of-K encoding
E_Q_FAM1	설문 3-1:안과질환 가족력 질환	□,□,□ (중복응답,예 1,2,3) 1.녹내장 2.백내장 3.사시 4.안검하수 5.망막 6.기타 안과질환 888888.비해당 999999.무응답	1-of-K encoding 888888->0 999999->9
E_Q_RM	설문 4:수족냉증/편두통 유병여부	1.수족냉증 2.편두통 3.수족냉증과편두통 4.없음 8.비해당(12 세미만) 9.무응답	1-of-K encoding
E_Q_EXC	설문 5:1 년간 안과검사 여부	1.예 2.아니오 8.비해당(12 세이상) 9.무응답	1-of-K encoding
E_Q_EXC1	설문 5-1:1 년간 안과검사 기관	□,□,□ (중복응답,예 1,2,3) 1.건강보험공단 신체검사 소아과 2.건강보험공단 신체검사로 안과 3.안경점 4.안과 5.학교 6.기타 888888.비해당 999999.무응답	1-of-K encoding 888888->0 999999->9

변수명	변수명	상세	정규화
E_Dmh	당뇨병 과거력	1.있다 2.없거나 모른다 8.비해당(19 세미만) 9.무응답	1-of-K encoding
E_VS_SQ	사시_유병여부	0.없음 1.있음	1-of-K encoding
E_VS_MRD	안검하수 유병여부	0.없음 1.있음	1-of-K encoding
E_Tr_y	백내장_백내장유무_우안	1.있다 2.없다 3.인공수정체안 4.무수정체안 8.비해당(19 세미만) 9.무응답(미검진)	1-of-K encoding
E_Tl_y	백내장_백내장유무_좌안	1.있다 2.없다 3.인공수정체안 4.무수정체안 8.비해당(19 세미만) 9.무응답(미검진)	1-of-K encoding
E_VS_TY	백내장 유병여부	0.없음 1.있음	1-of-K encoding
E_VS_WY	군날개 유병여부	0.없음 1.있음	1-of-K encoding
AC1_yr	1 년간 손상 발생여부	1.예 2.아니오 9.모름	1-of-K encoding
AC3_3_01~03	손상 1~3 기전	1.운수사고 2.추락/미끄러짐 3.부딪힘 4.열상/자상/절단/관통상 5.화상 6.질식 7.익수 8.중독 9.기타 99.모름	1-of-K encoding
AC8_1_01~03	손상 1~3 치료	1.응급실 2.병의원 외래만 이용 3.병의원 병실입원 4.기타 9.모름	1-of-K encoding
AC3_4_01~03	손상 1~3 의도성	1.불의의 사고 2.의도적 자해 3.다른 사람의 폭력 9.모름	1-of-K encoding
AC8_2w_01~03	손상 1~3 와병여부	1.예 2.아니오 9.모름	1-of-K encoding
ij_expr	연간손상경험여부	1.최근 1 년간 사고중독발생 경험 0.최근 1 년간 사고중독발생 미경험	1-of-K encoding

변수명	변수명	상세	정규화
MH1_yr	1년간 입원이용 여부	1.예 2.아니오 9.모름	1-of-K encoding
MH3_1_01~06	입원 1~6 이용이유	1.질병 2.사고중독 3.기타 9.모름	1-of-K encoding
MO1_wk	2주간 외래이용 여부	1.예 2.아니오 9.모름	1-of-K encoding
MO3_2_01~14	외래 1~14 이용이유	1.질병 2.사고중독 3.기타 9.모름	1-of-K encoding
MP1_wk	2주간 약국이용 여부	1.예 2.아니오 9.모름	1-of-K encoding
MP4_2_01~8	약국 1~8 구매품 -처방약	1.예 0.아니오	1-of-K encoding
MP4_3_01~8	약국 1~8 구매품 -비처방 일반의약품	1.예 0.아니오	1-of-K encoding
MP4_5_01~8	약국 1~8 구매품 -의약외 품	1.예 0.아니오	1-of-K encoding
MP4_6_01~8	약국 1~8 구매품 -기타	1.예 0.아니오	1-of-K encoding
O_PROS_U	상약 보철물상태	0.보철물 없음 1.고정성가공의치 한 개 2.고정성가공의치 두 개 이상 3.국소의치만 있음 4.고정성가공의치와 국소의치공존 5.총의치	1-of-K encoding
O_PROS_L	하악 보철물상태	0.보철물 없음 1.고정성가공의치 한 개 2.고정성가공의치 두 개 이상 3.국소의치만 있음 4.고정성가공의치와 국소의치공존 5.총의치	1-of-K encoding
O_IMP_U	상악 임플란트존재	0.없음 1.1 개 이상 존재	1-of-K encoding
O_IMP_L	하악 임플란트존재	0.없음 1.1 개 이상 존재	1-of-K encoding
O_DENT_U	상악 국소의치 및 총의치 필요	0.없음 1.국소의치 필요 2.총의치 필요	1-of-K encoding
O_DENT_L	하악 국소의치 및 총의치 필요	0.없음 1.국소의치 필요 2.총의치 필요	1-of-K encoding

Appendix 5. MATLAB code of DT, RF, SVM algorithm

5.1 DT algorithm

5.1.1 eq5d_step4_dt_viewtree.m

```
1 clear all;
2 clc;
3 nClass = 2;
4 kfold = 10;
5
6 load EQ5D_v22.mat;
7
8 [nRows nCols] = size(X);
9
10 c=zeros(1, nCols); c(135:end)=1; c=logical(c);
11
12 COST = [ 0 10; 1 0;];
13
14 tree = ClassificationTree.fit(X, Y, 'Cost', COST,
    'CategoricalPredictors', c, 'ScoreTransform', 'logit');
15
16 fn = sprintf('eq5dres_dt_v22_view_final.mat');
17 save(fn, 'tree');
18
19 view(tree);
20 view(tree, 'mode', 'graph');
```

5.1.2 eq5d_step5_dt_10cv.m

```
1 clear all;
2 clc;
3 nClass = 2;
4 kfold = 10;
5 load EQ5D_v22.mat;
6
7 cm = zeros(nClass, nClass, kfold);
8
9 [nRows nCols] = size(X);
10 c=zeros(1, nCols); c(135:end)=1; c=logical(c);
11 cv = cvpartition(Y,'kfold',kfold);
12
13 COST = [ 0 10; 1 0;];
14
15 for k=1:kfold
16     trIdx = cv.training(k);
17     tsIdx = cv.test(k);
18
19     trX = X(trIdx,:);
20     trY = Y(trIdx,:);
21
22     tsX = X(tsIdx,:);
23     tsY = Y(tsIdx,:);
24
25     tree = ClassificationTree.fit(trX, trY, 'Cost', COST,
        'CategoricalPredictors', c, 'ScoreTransform', 'logit');
```

```

26     yp = predict(tree, tsX);
27
28     cm(:, :, k) = confusionmat(tsY, yp);
29     cm(:, :, k) % Just for Viewing
30 end;
31
32 %%
33 % Perf Score
34 perfscore = zeros(kfold, 8);
35 for i=1:10
36     scm = cm(:, :, i);
37     TP = scm(1,1);
38     FN = scm(1,2);
39     FP = scm(2,1);
40     TN = scm(2,2);
41     TPR = TP / ( TP + FN);
42     FPR = FP / ( FP + TN);
43     Sens = TPR;
44     Spec = TN / ( FP + TN);
45     Prec = TP / ( TP + FP);
46     Recall = TPR;
47     Acc = (TP+ TN) / ( TP + FN + FP + TN);
48     Fsco = 2*(Prec*Recall)/(Prec+ Recall);
49
50     perfscore(i, 1) = TPR;
51     perfscore(i, 2) = FPR;
52     perfscore(i, 3) = Spec;
53     perfscore(i, 4) = Sens;

```

```

54     perfscore(i, 5) = Prec;
55     perfscore(i, 6) = Recall;
56     perfscore(i, 7) = Acc;
57     perfscore(i, 8) = Fsco;
58 end;
59
60 figure;
61 boxplot(perfscore(:,1:8));
62 set(gca,'XTick',1:8)
63 set(gca,'XTickLabel',{'TRP', 'FPR', 'Specificity', 'Sensitivity',
    'Precision', 'Recall','Accuracy','F-Score'})
64 title('Performance');
65 %xlabel('Perf Measure');
66 ylabel('Score');
67
68 %%
69 %
70 fn = sprintf('eq5dres_dt_v22_10cv_final.mat');
71 save(fn, 'cm', 'perfscore');

```

5.2 RF algorithm

5.2.1 eq5d_step1_rf_find_ntree_10cv.m

```
1 clear all; clc;
2 load EQ5D_v22.mat;
3 kfold = 10; nClass = 2; maxNTree = 200;
4 Y = Y;
5 X = X;
6 COST = [ 0 10; 1 0;];
7 cv = cvpartition(Y,'kfold',kfold);
8 perfscore = zeros(maxNTree, 12);
9
10 for NTree=3:5
11     icm = zeros(nClass, nClass);
12     cm = zeros(nClass, nClass);
13     parfor k=1:kfold
14         trIdx = cv.training(k);    tsIdx = cv.test(k);
15         trX = X(trIdx,:);    trY = Y(trIdx,:);
16         tsX = X(tsIdx,:);    tsY = Y(tsIdx,:);
17         %rfOptions = statset('UseParallel','always');
18         bt = TreeBagger(NTree, trX, trY, 'Cost', COST,
19             'NVarToSample', 'all', 'method','classification', 'oobvarimp',
20             'on');%,'options', rfOptions);
19         yp = predict(bt, tsX);    yp = str2double(cellstr(yp));
20         cm = confusionmat(tsY, yp);
21         icm = icm + cm;
22     end;
23     NTree
```

```

24   icm
25   TP = icm(1,1);
26   FN = icm(1,2);
27   FP = icm(2,1);
28   TN = icm(2,2);
29   TPR = TP / ( TP + FN);
30   FPR = FP / ( FP + TN);
31   Sens = TPR;
32   Spec = TN / ( FP + TN);
33   Prec = TP / ( TP + FP);
34   ReCa = TPR;
35   Acc = (TP+ TN) / ( TP + FN + FP + TN)
36   Fsco = 2*TP / (2*TP + FP + FN)
37
38   perfscore(NTree, 1) = TP;
39   perfscore(NTree, 2) = FN;
40   perfscore(NTree, 3) = FP;
41   perfscore(NTree, 4) = TN;
42   perfscore(NTree, 5) = TPR;
43   perfscore(NTree, 6) = FPR;
44   perfscore(NTree, 7) = Sens;
45   perfscore(NTree, 8) = Spec;
46   perfscore(NTree, 9) = Prec;
47   perfscore(NTree, 10) = ReCa;
48   perfscore(NTree, 11) = Acc;
49   perfscore(NTree, 12) = Fsco;
50 end;
51

```

```

52 %View As Stairs Graph
53 [value index]=max(perfscore(:,end))
54 stairs(perfscore(:,end));
55 title('F-Score v.s # of trees');
56 xlabel('# of trees');
57 ylabel('F-Score');
58 hold on
59 stem(4, 0.3465, 'or')
60 hold off
61
62
63 fn=sprintf('eq5dres_rf_v22_max200nt_perfscore_10cv_final.mat');
64 save(fn, 'perfscore');

```

5.2.2 eq5d_step2_rf_FeatureSelection.m

```

1 clear all; clc;
2
3 NTree = 4;
4
5 load EQ5D_v22.mat;
6
7 Y = Y;
8 X = X;
9 COST = [ 0 10; 1 0;];
10
11 % Random Forest for feature selection
12 rfOptions = statset('UseParallel','always');

```

```

13  bt = TreeBagger(NTree, X, Y, 'Cost', COST, 'NVarToSample', 'all',
'method','classification', 'oobvarimp', 'on','options', rfOptions);
14
15  bar(bt.OOBPermutedVarDeltaError);
16  title('Feature Importance');
17  xlabel('X Variable Index');
18  ylabel('Weight');
19
20  fn=sprintf('eq5dres_rf_v22_featureselection_final.mat')
21  save(fn)

```

5.2.3 eq5d_step3_rf_10cv.m

```

1  clear all;
2  clc;
3
4  NTree = 4;
5
6  nClass = 2;
7  kfold = 10;
8
9  load eq5dres_rf_v22_featureselection_final.mat
10  fi = bt.OOBPermutedVarDeltaError;
11  [fir fic] = find(fi > 0);
12  fiv = fi(fic);
13  [v si]=sort(fiv, 'descend');
14  sfic = fic(si);
15

```



```

16 load EQ5D_v22.mat;
17 X = X(:, fic);
18
19 cm = zeros(nClass, nClass, kfold);
20 cv = cvpartition(Y,'kfold',kfold);
21
22 COST = [ 0 10; 1 0;];
23 for k=1:kfold
24     trIdx = cv.training(k);
25     tsIdx = cv.test(k);
26
27     trX = X(trIdx,:);
28     trY = Y(trIdx,:);
29
30     tsX = X(tsIdx,:);
31     tsY = Y(tsIdx,:);
32
33     bt = TreeBagger(NTree, trX, trY, 'Cost', COST,
        'NVarToSample', 'all', 'method','classification', 'oobvarimp',
        'on');
34     yp = predict(bt, tsX);
35     yp = str2double(cellstr(yp));
36     cm(:, :, k) = confusionmat(tsY, yp);
37     cm(:, :, k) % Just for Viewing
38 end;
39
40
41 %%

```

```

42 % Perf Score
43 perfscore = zeros(kfold, 8);
44 for i=1:10
45     scm = cm(:, :, i);
46     TP = scm(1,1);
47     FN = scm(1,2);
48     FP = scm(2,1);
49     TN = scm(2,2);
50     TPR = TP / ( TP + FN);
51     FPR = FP / ( FP + TN);
52     Sens = TPR;
53     Spec = TN / ( FP + TN);
54     Prec = TP / ( TP + FP);
55     Recall = TPR;
56     Acc = (TP+ TN) / ( TP + FN + FP + TN);
57     Fsco = 2*(Prec*Recall)/(Prec+ Recall);
58
59     perfscore(i, 1) = TPR;
60     perfscore(i, 2) = FPR;
61     perfscore(i, 3) = Spec;
62     perfscore(i, 4) = Sens;
63     perfscore(i, 5) = Prec;
64     perfscore(i, 6) = Recall;
65     perfscore(i, 7) = Acc;
66     perfscore(i, 8) = Fsco;
67 end;
68
69 figure;

```

```
70 boxplot(perfscore(:,1:8));
71 set(gca,'XTick',1:8)
72     set(gca,'XTickLabel',{'TRP', 'FPR', 'Specificity', 'Sensitivity',
'Precision','Recall','Accuracy','F-Score'})
73 title('Performance');
74 xlabel('Perf Measure');
75 ylabel('Score');
76
77 fn = sprintf('eq5dres_rf_v22_10cv_final.mat');
78 save(fn, 'cm', 'perfscore');
```

5.3 SVM algorithm

5.3.1 eq5d_step6_svm_paramsearch.m

```
1 clear all;
2 clc;
3 nClass = 2;
4 kfold = 10;
5
6 load eq5dres_rf_v22_featuresselection_final.mat
7 fi = bt.OOBPermutedVarDeltaError;
8 [fir fic] = find(fi ~= 0);
9 fiv = fi(fic);
10 [v si]=sort(fiv, 'descend');
11 sfic = fic(si); % Sorted Feature Index
12
13 load EQ5D_v22.mat;
14 Y2 = Y;
15 XX2 = X(:, sfic);
16
17 cv = cvpartition(Y2,'kfold',kfold);
18 c=-3:0.2:3;
19 g=-6:0.2:0;
20 nVars = length(sfic);
21
22 nc=length(c);
23 ng=length(g);
24
25 for p=1:nVars
```

```

26     perfmat=zeros(nc,ng);
27
28     X2=XX2(:,1:p);
29     X2 = bsxfun(@rdivide, bsxfun(@minus, X2, min(X2)),
30               abs(max(X2)-min(X2)));
31     X2(:,isnan(X2(1,:)))=[];
32
33     for i=1:nc % for C
34         for j=1:ng % for Gamma
35             cm = zeros(nClass, nClass);
36             icm = zeros(nClass, nClass);
37             svmparam = sprintf('-c %f -g %f -w0 10',
38                               exp(c(i)), exp(g(j)))
39             parfor k=1:kfold
40                 trIdx = cv.training(k);
41                 tsIdx = cv.test(k);
42                 trX = X2(trIdx,:);
43                 trY = Y2(trIdx,:);
44                 tsX = X2(tsIdx,:);
45                 tsY = Y2(tsIdx,:);
46                 svm = svmtrain(trY, trX, svmparam);
47                 yp = svmpredict(tsY, tsX, svm);
48                 cm = confusionmat(tsY, yp)
49                 icm(:, :,k) = cm;
50             end;
51         scm=sum(icm, 3);
52     scm
53     TP = scm(1,1);

```

```

52     FN = scm(1,2);
53     FP = scm(2,1);
54     TN = scm(2,2);
55     TPR = TP / ( TP + FN);
56     FPR = FP / ( FP + TN);
57     Sens = TPR;
58     Spec = TN / ( FP + TN);
59     Prec = TP / ( TP + FP);
60     Recall = TPR;
61     Acc = (TP+ TN) / ( TP + FN + FP + TN);
62     Fsco = 2*(Prec*Recall)/(Prec+ Recall)
63     perfmat(i, j) = Fsco;
64     end;
65 end;
66 fn = sprintf('eq5dres_svm_v22_param_perfmat_%d.mat', p);
67 save(fn, 'perfmat');
68 end;

```

5.3.2 eq5d_step7_svm_param_view.m

```

1 clear all;
2 clc;
3 nClass = 2;
4 kfold = 10;
5
6 load eq5dres_rf_v22_featuresselection_final.mat
7 fi = bt.OOBPermutedVarDeltaError;
8 [fir fic] = find(fi ~= 0);

```

```

9  fiv = fi(fic);
10 [v si]=sort(fiv, 'descend');
11 sfic = fic(si); % Sorted Feature Index
12
13 nVars = 64%length(sfic);
14
15 maxperf = zeros(nVars, 1);
16
17 for p=1:nVars; % do exp from rank #1 to #305 incrementally
18     fn = sprintf('eq5dres_svm_param_performat_%d.mat', p);
19     load(fn, 'performat');
20     maxperf(p) = max(max(performat));
21     max(max(performat));
22 end;
23
24 [c nFS] = max(maxperf)
25
26 fn = sprintf('eq5dres_svm_param_performat_%d.mat', nFS);
27 load(fn, 'performat');
28
29 [x y] = max(performat); % need to find max C G
30 [i j] = max(x);
31 i=y(j);
32
33 c=-3:0.2:3;
34 g=-6:0.2:0;
35 Cval=exp(c(i))
36 Gval=exp(g(j))

```

```

37 i
38 j
39
40 figure;
41 stairs(maxperf);
42 title('Performance per # of Features in SVM');
43 xlabel('# of Features in SVM (Order by Weight)');
44 ylabel('F-Score');
45
46 figure;
47 contourf(perfmat);
48 title('Performance Contour with C vs G in SVM');
49 ylabel('C Value : exp(-3:0.2:3)');
50 xlabel('G Value : exp(-6:0.2:0)');
51 % hold on
52 % plot(j,i, 'ob')
53 % hold off
54
55 figure; surf(perfmat);
56 title('Performance 3D Plot with C vs G in SVM');
57 ylabel('G Value : exp(-6:0.2:0)');
58 xlabel('C Value : exp(-3:0.2:3)');
59 zlabel('F-Score');

```

5.3.3 eq5d_step8_svm_10cv.m

```

1 clear all; clc;
2

```



```

3 load eq5dres_rf_v22_featuresselection_final.mat
4 fi = bt.OOBPermutedVarDeltaError;
5 [fir fic] = find(fi>0);
6 fiv = fi(fic);
7 [v si]=sort(fiv, 'descend');
8 sfic = fic(si); % Sorted Feature Index
9 nFS=32; % # of best Feature
10 load EQ5D_v22.mat;
11
12 Y2 = Y;
13 X2 = X(:, sfic(1:nFS));
14
15 X2 = bsxfun(@rdivide, bsxfun(@minus, X2, abs(min(X2))),
abs(max(X2)-min(X2)));
16 X2(:,isnan(X2(1,:)))=[];
17
18 nClass = 2; kfold = 10;
19 %svmparam = '-c 0.367879441171442 -g 0.074273578214334
-w0 10'
20 %svmparam = '-c 2.72 -g 0.30 -w0 10'
21 svmparam = '-c 0.819 -g 0.3012 -w0 10'
22
23 cv = cvpartition(Y,'kfold',kfold);
24 cm = zeros(nClass, nClass, kfold);
25
26 for k=1:kfold
27     trIdx = cv.training(k);
28     tsIdx = cv.test(k);

```

```

29     trX = X2(trIdx,:);
30     trY = Y2(trIdx,:);
31     tsX = X2(tsIdx,:);
32     tsY = Y2(tsIdx,:);
33
34     svm = svmtrain(trY, trX, svmparam);
35     yp = svmpredict(tsY, tsX, svm);
36
37     cm(:, :, k) = confusionmat(tsY, yp);
38     cm(:, :, k)
39 end;
40
41 %%
42 % Perf Score
43 perfscore = zeros(kfold, 8);
44 for i=1:10
45     scm = cm(:, :, i);
46     TP = scm(1,1);
47     FN = scm(1,2);
48     FP = scm(2,1);
49     TN = scm(2,2);
50     TPR = TP / ( TP + FN);
51     FPR = FP / ( FP + TN);
52     Sens = TPR;
53     Spec = TN / ( FP + TN);
54     Prec = TP / ( TP + FP);
55     Recall = TPR;
56     Acc = (TP+ TN) / ( TP + FN + FP + TN)

```

```

57     Fsco = 2*(Prec*Recall)/(Prec+ Recall)
58
59     perfscore(i, 1) = TPR;
60     perfscore(i, 2) = FPR;
61     perfscore(i, 3) = Spec;
62     perfscore(i, 4) = Sens;
63     perfscore(i, 5) = Prec;
64     perfscore(i, 6) = Recall;
65     perfscore(i, 7) = Acc;
66     perfscore(i, 8) = Fsco;
67 end;
68 mean(perfscore(:, 7))
69 mean(perfscore(:, 8))
70 figure;
71 boxplot(perfscore(:,1:8));
72 set(gca,'XTick',1:8)
73 set(gca,'XTickLabel',{'TRP', 'FPR', 'Specificity', 'Sensitivity',
    'Precision','Recall','Accuracy','F-Score'})
74 title('Performance');
75 xlabel('Perf Measure');
76 ylabel('Score');
77
78 fn = sprintf('eq5dres_svm_v22_10cv_final.mat');
79 save(fn, 'cm', 'perfscore');

```

Abstract

Analysis of Influencing Factors and Development of Prediction Model for the Health related Quality of Life in Elderly with Chronic Diseases

Soo-Kyoung Lee

Healthcare Management and Informatics

The Graduate School

Seoul National University

The aging of the population and the consequent increase of chronic diseases degrade the health related quality of life (HRQoL) of the elderly with chronic diseases. Therefore, there is a need to provide effective intervention on HRQoL improved by identifying the factors affecting HRQoL of the elderly with chronic diseases and development of prediction model for reducing HRQoL of the risk group. The purpose of this study was to attempt comprehensive analysis of the factors affecting HRQoL of the elderly with chronic diseases and develop a prediction model of HRQoL risk group that required interventions.

Because it is required considerable effort and expertise for multi-dimensional analysis on the complex concept of HRQoL, It needs to be applied in a variety of ways that can give support these complex analytical processing automatically. Especially, by applying

the latest machine learning (ML) techniques that facilitate the analysis and prediction of large volumes of data, we examined the possibility of application of ML techniques for analysis and prediction of HRQoL of the elderly with chronic diseases and the consequent problems and solutions.

In this study, this material was used for the 4th and 5th, which was conducted from 2008 to 2010, the National Health and Nutrition Examination Survey (KNHANES). This survey sponsored by the Centers for Disease Control and Prevention and the results obtained from the nationwide survey of year-round (approximately 10,000 people randomly extracted) consisting of health interview, screening and nutrition examination survey. The final 716 cases were selected for this study. For the data, the elderly over the age of 65 with a chronic disease were 2,991 cases in entire 18,546 cases and then the missing values and outlier were removed in 2,991 cases.

For data analysis and development of prediction model, we used SPSS (Statistical Package for the Social Sciences) version 20.0 and MATLAB (Matrix Laboratory) Revised for Version 7.14 Release 2012a by using descriptive statistics, X^2 -test, stepwise logistic regression analysis (SLR). We carried out to apply the algorithms as follows: ML technique of decision tree (DT), random forest (RF), support vector machine (SVM) methods. The HRQoL was measured by the EQ-5D (European Quality of Life-5 Dimensions). Based on the EQ-5D index 0.678, the cases were classified into over 0.678 group (n=648) or below 0.678 group (n=68).

The affecting factors of HRQoL in the elderly with chronic diseases showed statistically significance were 'income', 'chronic disease type', 'depression', 'discomfort', 'perceived health status'. The results of the

application ML for prediction of HRQoL of the elderly with chronic diseases have been identified as additional important variables that 'osteoarthritis duration', 'prevalence of osteoporosis', 'stroke timing of diagnosis', 'state of the mandibular prosthesis', '1year suicidal ideation' etc. For validation of a predictive model developed, we applied the 10-fold cross-validation method. Accuracy, F-score values were calculated by the each model in order to evaluate the performance. In the results, SLR analysis achieved Accuracy=0.93, F-score=0.49 as the best performance was shown, then the SVM (0.90, 0.51), RF (0.87, 0.33), DT (0.82, 0.23) respectively.

As a result, in applying machine learning techniques for such a mixed type of data KNHANES, through the process of variables selection and data normalization was able to improve the performance and efficiency of the prediction model. In the future, ongoing research and verification process is required and related research will be able to provide specific guidelines on how to select variables and normalize the data on similar types of data utilization.

Through this study, we will be able to provide a variety of basic materials that are required in personalized health management strategies and help developing of interventions programs to improve HRQoL for elderly people with chronic diseases. As seen in this study, to derive valuable new knowledge through the data repositioning in the medical field, we hope that such an approach would be applied more vigorously.

Keywords : elderly, chronic diseases, health related quality of life,
prediction model, machine learning

Student Number : 2008-31047

감사의 글

무모하게 시작했던 늦은 학업의 길이 졸업이라는 마침표와 동시에 제게는 인생의 터닝포인트가 되어 다시금 새로운 길이 되었습니다. 이제까지 꿈꾸어왔던 소박하지만 절실했던 꿈너머 꿈이 이루어지기를 기도하며 첫걸음을 내딛고자 합니다. 약함 가운데 소망을 주시고, 의미 있는 삶을 살도록 인도해주시는 하나님, 저는 축복받은 사람입니다. 앞으로도 끊임없이 노력하고, 최선을 다하는 모습으로 살아가겠습니다. 지금까지 저를 아껴주시고, 격려해주신 모든 분들께 작은 결실인 논문을 감사의 마음으로 드립니다.

먼저 제자로 받아주시고 끝까지 학업에 정진하며 나아갈 수 있도록 진심 어린 조언과 지도로 계속 발전하는 학자로서의 여정을 이끌어주신 김홍기 교수님께 고개 숙여 깊이 감사드립니다. 소중한 인연으로 지도교수님이 되어주셔서, 변함없는 따뜻한 격려와 정성어린 세심한 지도로 무사히 학업을 마칠 수 있도록 애써주신 Black Lady Hawk 김정은 교수님께도 깊은 감사를 드립니다. 대학동기이자 친구인 저에게 든든한 후원자가 되어주고, 학문적인 성장을 위해 세심하고 명확한 지도를 아끼지 않았던 손연정 교수님께도 깊은 신뢰와 감사의 마음을 전하고 싶습니다.

바쁜 와중에도 기꺼이 심사위원장님을 맡아주시고, 연구의 중심과 방향이 흔들리지 않도록 핵심적인 내용을 지도해주셨던 이재일 교수님과 새로운 학문분야에 대하여 함께 고민해주시고, 지혜롭게 학업과 육아를 병행할 수 있는 조언까지 아끼지 않으셨던 강보영 교수님께 진심으로 감사를 드립니다. 평생을 의료정보학의 발전을 위해 애쓰신 김명기 교수님과 변함없는 열정으로 학문과 봉사의 삶을 살아가시는 채영문 교수님께도 무한한 존경과 깊은 감사의 마음을 전하고 싶습니다.

4년 반이라는 길고도 짧은 기간 동안 울고 웃으며 함께 지냈던 의생명지식공학연구실의 모든 식구들, 저의 졸업을 위해 너무나 큰 힘이 되어주신 정상원 선생님과 직장으로 복귀하시는 날까지 실험을 위해 밤을 새며 애써 주신 조현성 선생님, 그리고 다시 학업의 길을 시작할 수 있도록 도와주고, 포기하지 않도록 끝까지 응원해준 유정숙 선생님, 연건캠퍼스에서 학업을 하면서 함께 고민을 나누며 힘이 되어준 이명경 선생님, 늘 친동생같이 아껴주고 결혼의 인연까지 선물해준 현주언니, 서영이를 사랑으로 보살펴준 시후네 가족들에게도 깊은 감사의 마음을 전합니다.

변함없는 지지와 응원을 해준 사랑하는 엄마와 가족들, 부족한 며느리를 아껴주시고 격려해주신 아버님과 어머님에게 진심으로 감사드립니다. 힘든 상황을 이겨나갈 수 있도록 아낌없이 지원해준 사랑하는 남편과 건강하게 잘 자라준 서영이에게 정말 고맙다는 말을 전하고 싶습니다. 마지막으로 늦은 나이에 공부라는 새로운 모험을 시작했을 때 유일하게 잘했다고 칭찬해주셨고, 살아계셨다면 저의 졸업을 그 누구보다도 기뻐해주셨을 너무나도 그리운 아빠에게 이 논문을 바칩니다.

2013년 1월 창경궁을 바라보며
이 수 경